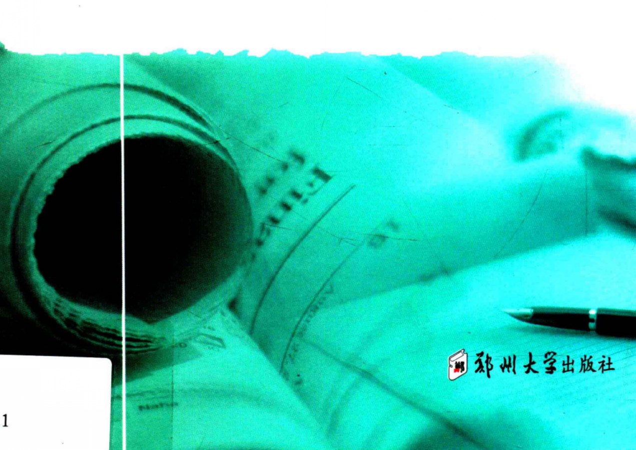
青年科技创新人才学术文库

GIS支持的土壤重金属污染评价与分析

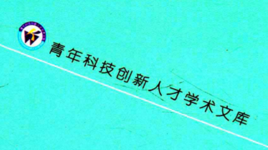
李　向　李玲玲　著



策划编辑 刘 开
责任编辑 刘 开
责任校对 李 蕊
封面设计 张 庆
版式设计 叶 紫

GIS支持的土壤重金属污染评价与分析

GIS ZHICHI DE TURANG ZHONGJINSHU WURAN PINGJIA YU FENXI





青年科技创新人才学术文库

GIS支持的土壤重金属污染评价与分析

李 向 李玲玲 著

郑州大学出版社
郑 州

图书在版编目(CIP)数据

GIS支持的土壤重金属污染评价与分析/李向,李玲玲著.—郑州:郑州大学出版社,2012.6

ISBN 978-7-5645-0844-9

Ⅰ.①G…　Ⅱ.①李…　②李…　Ⅲ.①地理信息系统-应用-土壤污染-重金属污染-污染物分析　Ⅳ.①X53

中国版本图书馆CIP数据核字(2012)第095201号

郑州大学出版社出版发行
郑州市大学路40号　　邮政编码:450052
出版人:王　锋　　发行部电话:0371-66966070
全国新华书店经销
郑州文华印务有限公司印制
开本:787 mm×1 092 mm　1/16
印张:9.25
字数:222千字
版次:2012年6月第1版　　印次:2012年6月第1次印刷

书号:ISBN 978-7-5645-0844-9　　定价:30.00元

前 言

重金属的污染与危害已成为人类所面临的重要环境问题之一，尤其是土壤中重金属污染。这是由于随着现代城市化进程和工业生产的迅速发展，工业“三废”排放量逐年增加，含重金属的化肥、杀虫剂和除草剂被大量使用，使土壤受到不同程度的重金属污染。重金属污染会影响土壤的正常功能，严重威胁着其上生长的植物农产品的安全，从而对环境安全造成威胁。

随着环境科学、土壤科学的发展，土壤重金属污染研究越来越受到人们的广泛关注。研究土壤、植物、大气、水等环境要素中重金属循环和累积的过程及规律，不仅具有重要的理论价值，而且对防治重金属污染、保障人体健康具有实际指导意义。

在国土资源大调查项目“不同景观城市的生态地球化学环境调查与风险评估方法技术研究（项目编号1212010660406）”的支持下，国家地质实验测试中心及有关单位开展了“典型地区土壤地球化学动态监测及预警预报”的研究。

国家地质实验测试中心刘晓端研究员所带领的课题组在城市环境地球化学调查技术方法研究方面作出了重要贡献。该研究是生态地球化学系统研究的一部分，所研究的城市生态系统的主要环境要素包括地质背景、水文地质、地理环境、土壤母质、社会环境等。在徐清博士的推荐下，郑州航空工业管理学院李玲玲博士所带领的“智能信息处理”研究团队参加了课题组的部分研究工作，利用遥感、地理信息系统（GIS）及生态地球化学技术对工业城市大气、土壤、水环境进行了监测，完成了“土壤地球化学环境重金属污染动态监测及预测/预警系统”的研究和开发。

本书介绍了“土壤地球化学环境重金属污染动态监测及预测/预警系统”研究工作的成果，主要内容包括土壤重金属污染评价技术、GIS 支持的土壤重金属污染评价、土壤重金属污染数据的空间分析、多种土壤重金属污染评价模型、土壤重金属污染预测模型、典型应用介绍等。

本书将 GIS 技术与土壤重金属污染研究相结合，实现了评价及预测模型。对于环境工程及评价人员而言，提供了方便有效的环境质量评价模型和软件平

台。对于 GIS 开发人员而言,从工程应用开发的角度介绍了以特定行业应用为导向的 GIS 软件二次开发方法。

本书在撰写过程中引用了大量的国内外相关领域的最新成果和资料,在此向这些专家、学者致以衷心的感谢。本书第一章至第六章、第十章由李向撰写;第七章至第九章、第十一章、第十二章由李玲玲撰写。最后,由李玲玲对全书进行审定。书稿的写作虽然已经数易其稿,但是由于作者能力和时间的有限,不足之处恳请广大读者批评指正。

李　向

2012. 3

目　录

第1章

土壤重金属污染评价技术

土壤是连续覆盖在地球陆地表面具有肥力的疏松物质，是一种随着气候、生物、母质、地形和时间等因素发生变化的自然资源。典型的土壤随深度呈现不同的层次，通常将地面植被或覆盖物去除后，深度为0~20 cm的土壤称为表层。表层也称淋溶层，是土壤中生物作用最活跃的一层，有机质大部分在这一层，金属离子和黏土颗粒在此层的淋溶过程也最明显。

土壤污染是指人类活动产生的污染物进入土壤并积累到一定程度，超出了土壤自我吸附、分解、迁移、转化的能力，进而引起土壤自然功能失调、土壤质量恶化的现象。土壤污染物质的主要来源是过度耕作生产、工业废水、生活污水、污染气体以及固体废弃物。土壤污染具有隐蔽性、滞后性、累积性、持久性、难治理的特点。

土壤污染，尤其是土壤中重金属污染物的富集，将会对农作物、环境以及人体等造成很大的危害，例如导致种植物产品品质下降，造成严重的经济损失，生物食用后引发癌症及其他疾病，导致生态环境系统恶化。

1.1 重金属污染及其危害

日益严重的环境压力已经危及人们的生活质量和身体健康，尤其是重金属污染对人类生存及生态安全的威胁非常大。

1.1.1 重金属污染

重金属是指相对密度大于4或5的金属，约有45种。重金属一般以天然浓度广泛存在于自然界中，铜、铅、锌、铁、锰、镉、汞、钨、金、银等是较为常见的重金属。这些重金属中，锰、铜、锌等是生命活动所需要的微量元素，但是大部分如汞、铅、镉等并非生命活动所必需。特别要说明的是，所有重金属，只要超过一定浓度，都对生物体有毒害作用。由于人类对重金属的开采、冶炼、加工及商业制造活动日益增多，造成不少重金属如铜、铅、汞、镉等进入大气、水、土壤中，造成严重的生态环境污染。

重金属污染指由重金属或其化合物造成的环境污染，主要由采矿、废气排放、污水灌溉和使用重金属制品等人为因素所致。随着城市化进程的急速进展，各种产业竞相掠夺环境资源，这一趋势加剧了工业、生活废弃物的排放及城市污染。此外，人们生活消费形

式的多样化带来农业种植方式的巨大变化，建设用地逐年增加、耕地面积锐减导致农药过多使用和化肥偏施等掠夺土地资源的现象加重，不可避免地加重了城市环境的重金属污染。

1.1.2 重金属污染的危害

重金属污染存在于水体、土壤、大气和固体废物中。以各种化学状态或化学形态存在的重金属，在进入环境或生态系统后就会存留、积累和迁移，对环境质量和人类健康造成很大的危害。重金属污染的危害程度取决于重金属在环境、食品和生物体中存在的浓度和化学形态。

大气颗粒物中，重金属是重要成分之一，主要来源于机动车尾气的排放、金属的冶炼、矿山的开采、建筑材料生产所产生的气体和粉尘等。除汞以外，重金属基本上是以气溶胶的形态进入大气，城市大气中的重金属通过呼吸作用和皮肤吸收进入人体，可导致高血压、心脏病发病率的上升，其中钒含量的增加与某些癌症的死亡率有着相关性。此外，大气中的重金属还经过自然沉降和降水进入土壤。

水体中的重金属主要来源于工业生产的废水。即使水体中重金属浓度再小，也可在藻类和底泥中积累，被鱼和贝的体表吸附，产生食物链浓缩，从而对人类产生危害。例如，日本的水俣病和骨痛病，前者由汞污染引发，后者由镉污染引起。水俣病是因氢氧化钠制造工业排放的废水中含有汞，再经生物作用转变成有机汞所造成的；骨痛病是由炼锌工业和镀镉工业所排放的镉所致。

土壤中重金属累积的外源因子很多，包括化肥和农药的使用、污水灌溉、污泥、城市垃圾、农业和工业废弃物排放、大气沉降等。已有研究表明，汞、铅等在一些区域，特别是工业集中或矿业发达地区的富集趋势已十分明显。进入土壤-植物系统中的重金属，在土壤中移动性很小，不易随水淋滤，不为微生物降解，通过食物链进入人体后，潜在危害极大。例如，土壤中汞能在植物体内累积，对植物产生毒害，引起植物汞中毒，严重情况下引起叶子和幼蕾掉落。经由食物链传递的重金属将直接危害人体健康。汞化合物侵入人体后，经血液循环迅速弥散到全身各器官，引起肾脏损害。又如，土壤中过量的镉会对植物的生长发育产生明显的危害，能使植物叶片受到严重伤害，生长缓慢，植株矮小，根系受到抑制，造成作物产量降低。种植在这样土壤之上的大米、蔬菜受到污染，被人食用后，会引起骨痛病，严重损伤肾小管，出现糖尿病，造成肺部损害、心血管损害，甚至致癌、致畸、致突变。因此，应特别注意防止重金属对土壤的污染。

由于大气-水-土壤生态系统是一个循环体系，所以有关陆地生态系统重金属污染物循环迁移累积规律的研究已成为环境科学领域的热点问题。据相关报道，许多工业发达国家，大气沉降对土壤系统中重金属累积贡献率在各种外源输入因子中排在首位。例如，日本的汽车尾气排放的铅经大气扩散等过程进入环境中，造成地表铅的浓度显著提高，致使近代人体内铅的吸收量比原始人增加了约100倍，严重损害了人体健康。

世界卫生组织和美国疾病控制与预防中心的研究都显示，一旦重金属摄入过量，不仅会引发胃病、恶心、呕吐、腹泻等症状，同时也会引发血液、肺、肾、脑、胰腺以及骨骼等

综合性疾病。此外，它还可能使得人类的免疫系统和神经系统受到损害，直至死亡。尤其是儿童以及未出生的婴儿，受到的伤害就更大。特别需要指出，幼儿及青少年接触重金属所造成的健康危害极有可能持续到青少年及成年阶段。例如，儿童血铅中毒，其最主要的影响是造成孩子的智力发育及生长发育滞后或停止，这种危害是不可逆转的。血铅中毒的临床表现有食欲不振、胃疼、失眠、学习障碍、便秘、恶心、腹泻、疲劳、智商低下、贫血等。重金属铅对人体的危害还远不止于此，人体铅中毒后，大约10%的铅存在于血液中，虽然这部分铅可以通过排铅治疗取得较好的效果，但是90%的铅存在于骨骼系统，毒性持久。在我国，血铅超标事件已经引起政府和公众的极大关注。2011年5月，中国环境保护部公开通报了浙江湖州德清血铅超标事件查处情况。据通报，海久电池股份有限公司周边多名儿童及成人被检测出血铅超标，其中成人233人，儿童99人。

美国疾病控制与预防中心定期公布全国性的化学污染与人体健康报告，其中就包括各种重金属的健康危害及在人体中存在的剂量。我国于2011年6月启动“环境重点污染物健康危害的监测评价与控制专项”，由国家投入总经费2 685万元，其中关于重金属的项目投入为800万元，首次开展环境重金属检测技术标准化与人体负荷监测。该专项的目标是建立重金属污染与健康危害监测数据库。

综上所述，重金属污染已成为人类所面临的重要环境问题之一，尤其是土壤重金属污染。随着现代城市化进程的加快和工业生产的发展，工业“三废”排放量增加，含重金属的化肥、杀虫剂和除草剂被大量使用，使土壤受到不同程度的重金属污染。重金属污染会影响土壤的正常功能，严重威胁着其上生长的植物农产品的安全，对环境安全造成威胁。生态环境四大要素中的大气受到污染后，其悬浮物沉降是土壤中重金属的重要来源。水体中的重金属通过生态圈的物质循环，影响重金属在土壤、大气中的含量变化。植物生长带来重金属迁移，被人类食用而直接危害人体健康。土壤中重金属不断积累，重金属在粮食、蔬菜等农产品中大量富集，对人体的伤害极大。土壤中的重金属天然分布十分广泛，多具有不可逆性，不能降解，但可以转化为金属有机物或在有些生物中高倍富集，可能经多种途径进入人体，并在人体内蓄积毒性。因此研究土壤、植物、大气、水等环境要素中重金属循环和累积的过程及规律，不仅具有重要的理论价值，而且对防治重金属污染、保障人体健康具有实际指导意义。

1.1.3 重金属污染的国内外研究现状

国内的污染生态学(Pollution Ecology)和国外的环境保护生态学(Ecology for Environment Protection)，都把重金属作为环境中的一种污染物优先进行研究。重金属污染研究领域很多，主要研究内容有重金属元素形态分析、生物有效性分析、生态风险评价、植物修复、植物富集等。各方面的研究，都有相应的分析和评价方法。

在这些研究中，由于土壤是人类赖以生存的基本自然资源，土壤污染中的重金属污染尤其需要人们的关注，得到了深入和广泛的研究。就我国而言，“十二五”规划提出了“提高生态文明水平”的主题和“加快建设环境友好型社会”的主线，根据生态友好和可持续发展的战略需求，工作重点之一是解决当前国内水域重金属污染、土壤重金属污染、

大气重金属污染等问题。下面对国内外土壤重金属方面的研究现状进行简要概述。

1.1.3.1 对土壤重金属含量的研究

国内外已有研究中,对土壤重金属污染的研究主要侧重于自然过程和技术角度。从20世纪50年代开始,发达国家相继发生了由化学污染引起的严重危害事件,金属污染监测成为必要,重金属的分析技术应运而生。

对土壤重金属含量进行监测的方法很多,主要分为异位监测法和原位监测法。

土壤重金属异位监测法是比较传统的监测方法,其监测的时间相对较长,精度较高。其一般步骤是样品田间采样,然后进行实验室分析和数据处理。实验室分析方法分为化学分析法和仪器分析法两种方法。

土壤重金属原位监测是指在不影响被测物状态及周围环境的前提下,对目标物含量进行实时在线跟踪测定。其思路多为利用全球定位系统定位被测区域,依靠原位分析技术,结合遥感技术获取光谱信息,通过对光谱信息整合计算获得被测物含量,进而用地理信息系统技术确定重金属空间分布特征。在土壤重金属原位监测中,将土壤遥感、全球定位、空气传播、卫星遥感、光学侦测和修正(LIDAR)、地理信息系统和数据库管理系统、处理多维环境数据的计算力量、先进的多元统计和地理统计方法等多学科整合,逐步实现连续、高密度监测,方便了土壤特性在时间和空间分布上的准确映射。目前支持原位分析法的技术主要有三大类,即高光谱分析技术、环境磁学技术和生物量间接测定技术。

最早开始收集世界各地大量土壤分析数据的是Swaine,他在1955年首次概述了世界土壤微量元素的平均含量,同年苏联学者对岩石、土壤等物质中元素的丰度做了进一步研究。

1961年美国地质调查局(United States Geological Survey)开始对美国本土的土壤开展背景值的调查研究工作。1972年美国环境质量委员会(Council on Environmental Quality)给出的报告指出,美国平均每人每天的镉摄取量为0.02~0.1 mg/kg。1976年美国环境保护署(United States Environmental Protection Agency,负责维护自然环境和保护人类健康不受环境危害影响的独立行政机构)根据土壤阳离子交换量的不同,制定了重金属的最高容许含量。

1.1.3.2 对土壤重金属污染或环境质量评价方法的研究

环境质量评价就是依据一定评价标准和方法对一定的区域内的环境质量进行评定和预测。由于环境质量评价涉及评价指标的选择、评价方法的确定以及评价结果的展示等诸多方面,因此,被用于环境质量评价的数学方法相对较多,但仍没有一个全面、统一的评价指标体系和定量方法。

土壤环境质量评价属于环境质量评价的大范畴,在环境日益恶化的今天,具有非同一般的作用和意义,是当前环境研究的一项重要内容。

20世纪六七十年代,国外关于土壤重金属污染的评价取得了重要进展。1969年德国科学家Muller提出了地积累指数法评价,并在欧洲被广泛采用。1980年瑞典科学家Hakanson基于元素丰度和释放能力的原则,提出了潜在生态危害指数法,广泛地应用于分析评价重金属的潜在生态影响。

国内外学者针对不同地区已经开展了很多土壤环境调查及土壤环境质量评价方面的研究。例如，1998 年，澳大利亚科学家 Lottermoser 等对 New South Wales 东北部的 Gulf Creek 铜矿区进行环境地球化学调查，研究结果表明，该地区土壤中重金属 As、Cu、Pb 和 Zn 的含量明显高于背景值。2002 年，土耳其科学家 Sponzaa 和 Karaoglu 对 Aliaga 金属工业区的环境地球化学质量进行了调查，对该地区的土壤、水体、水系沉积物中多种重金属的含量都进行了调查研究，并使用污染指数法评价了该区域的土壤环境质量。

我国自 20 世纪 80 年代开始开展土壤污染监测和治理技术方面的研究工作，对重金属背景值也有较多研究。1982 年，国家科委将土壤背景值调查研究列入"六五"重点科技攻关课题，选择湘江谷地和松辽平原开展了土壤环境背景值研究。早期研究的区域范围仅限于几个重点城市。1986 年国家再次将这一课题列为"七五"重点科技攻关课题，调查范围则扩大到了除台湾省以外的全国各省区。1988 年，中国完成了全国土壤背景值的调查研究工作，出版了《环境背景值数据手册》、《中国土壤元素背景值》等专著。

随着科技的迅猛发展，很多研究对土壤环境质量进行了评价，各种评价方法层出不穷，各具特色。从评价方法学的角度来说，对这些评价方法可以进行归类，主要有：①数理统计法，这是早期的环境质量综合指数法，是采用特定的统计方法对原始数据鉴别的综合性指标法。土壤质量指数、污染物标准指数等都是具有代表性的数理统计法。②层次分析法，是将评价主体分成若干层次，再把同一层次的各因子以上一层因子为准则，进行比较，并计算出权重。由于该法较成熟，且较为科学，因此被广泛采用。③主成分分析法，就是采用最少量的综合指标，客观地确定其权重。对土壤有效态、有机质或某种重金属进行研究和评价，多用主成分分析法。④模糊数学法，主要是应用模糊综合评判模型，对环境质量进行综合评价。其优点是采用模糊的、不确定的、综合的指标来评价环境质量。⑤灰色关联法，被广泛应用于各领域。灰色关联法具有良好的评价环境质量等级的优越性。⑥BP 神经网络法，被认为具有运算速度快、容错能力和自学能力强等特点，该法不仅综合评价各个指标的作用，还特别表现了主要影响因素的作用，具有较好的评价效果。

从 20 世纪 80 年代末开始，我国对土壤重金属污染问题开始进行较为深入的研究。1989 年，吴玉燕等报道了沈阳张士灌区土壤 Cd 污染状况，揭开了中国土壤重金属污染调查的序幕。1992 年，夏增禄提出了中国主要土类的 Cu、Pb、Cr、As 4 种重金属元素的土壤临界含量。1995 年我国颁布了《中华人民共和国土壤环境质量标准》(GB 15618—1995)，为进行土壤环境质量评价提供了重要依据。1996 年，张民等研究了我国菜园土壤中某些重金属元素的含量与分布，为预测和防治菜园土壤的重金属污染提供了依据。其他相关研究也很多，特点是针对某一区域，侧重某种重金属的研究。例如，2003 年，郑袁明等以北京市为例，研究了 Cr 和 Ni 两种重金属的空间分布，并对区域的土壤环境质量做出评价。2004 年，王素琴对江苏典型农业区的土壤进行了本底调查，并使用单项污染指数法对该地区的土壤环境质量做出评价。相对来说，已有研究中对土壤重金属污染空间分布的研究较少。

1.1.3.3 对土壤环境质量标准制定的研究

在土壤环境质量标准制定及研究方面，许多国家的土壤标准采用"基于风险的标

准”。例如,1996 年,美国环保总署颁布了旨在保护人体健康的土壤筛选导则(Soil Screening Guidance,SSG),还颁布了旨在保护生态受体安全的土壤生态筛选导则(Ecological Soil Screening Guidance,ESSG),美国许多州都据此制定了各州的土壤质量指导值。2002 年,英国环境署颁布了考虑不同土地利用方式下人体健康暴露风险而制定的土壤质量指导值。2009 年,英国环境署修订后发布了最新的污染土地健康风险评估的技术方法。1994 年,荷兰研究提出了开展污染土壤健康风险评估的技术方法,探讨了人群对土壤污染的暴露途径及模型评估方法,并将该方法用于保护人体健康的土壤基准的制定,2008 年荷兰环境部修订印发了最新的污染土壤风险管理和修复技术文件。此外,加拿大、澳大利亚、法国、瑞典、日本、越南、丹麦等国,也都制定了各自国家的土壤质量保护值。

有专家指出,目前国际上对土壤质量指导值或者说标准的命名方式各不相同。例如,加拿大谓之土壤质量指导值,美国谓之土壤筛选值,英国谓之土壤指导值,澳大利亚谓之土壤调研值,荷兰谓之目标值和干预值等。各国命名中均未直接包括“环境”二字,但实际却涵盖了两方面内容:一方面,为了保护土壤的生态功能,指导值基于生态毒理数据制定;另一方面,为了保护人体暴露于土壤污染物的无显著健康风险,指导值基于人体的健康暴露风险评估制定。

鉴于问题的复杂性和一些学术争议,我国从 2006 年开始,投入大量人力财力,用很长的时间,获得了全国大范围的土壤污染基础数据,并正在对相关标准草案进行整理和修改,以期得到科学、客观的标准。

1.1.3.4 对土壤重金属修复技术的研究

土壤重金属污染的修复方法主要有物理化学法、化学修复法、微生物修复法和植物修复法等。

物理化学法主要通过化学固化、土壤淋洗等过程修复。化学固化即在土壤中加入固化剂,改变土壤的理化性质,使重金属被吸附在土壤中或者形成沉淀,从而一定程度上降低重金属的生物有效性,降低毒性。但是重金属在固化之后仍滞留在土壤中,并且土壤中必需的化学元素也因固化剂的作用发生沉淀,土壤性质很难恢复。土壤淋洗主要是用能提高重金属可溶性的试剂,如有机或无机态酸、碱、盐和螯合剂等,将土壤固相中的重金属转移到液相中从而修复污染土壤。

化学修复法主要包括化学改良和有机质改良,通过向土壤中添加一些改良剂,使土壤的 pH 值改变,从而减轻重金属污染程度。例如,石灰就是一种很好的改良剂。常用的改良剂还有磷酸盐、硅酸盐、海泡石等。

微生物对重金属的生物富集主要表现在胞外络合、沉淀以及胞内累积。一些微生物如动胶菌、蓝细菌、硫酸还原菌以及某些藻类,能够产生具有大量阴离子基团的胞外聚合物(如多糖、糖蛋白等),与重金属离子形成络合物,从而将其从土壤中有效去除。微生物通过对重金属进行生物氧化与还原、甲基化与去甲基化以及重金属的溶解度和有机络合配位降解转化重金属,改变其毒性,从而形成某些微生物对重金属的降解机制。中科院微生物所的研究人员在实验中发现,烟草头孢酶 F2 对汞量减少有明显作用。

植物修复法主要指使用植物使重金属固定、挥发及提取。植物在土壤重金属修复中有着重要的作用。耐重金属污染的植物及其根系微生物能够分泌物质使重金属在其根、

茎、叶、果实等部位吸附螯合形成沉淀，从而固定土壤中的污染物。Baker 等人在英国的 IACR-Rothamsted 试验站进行了植物积累的首次田间试验，结果显示，*T. carulercens* 在净化 Zn 污染土壤方面有极大的成效。目前，国际上报道的超积累植物已有 500 多种，例如苎麻对土壤中的汞具有良好的吸附作用，蜈蚣草对土壤中的砷有良好的吸收能力。植物修复技术是一项有前途的新技术，与其他修复技术相比其费用较低。

目前寻找治理重金属污染的生物修复途径，主要是从对土壤重金属超富集和非富集作物两个方向着手找寻，但从目前来看，国内科学家还没有找到一种经济、有效、适合大规模农田治理的科学修复模式，大多数土壤重金属治理模式还处于实验室摸索阶段。

所有的污染修复手段起到的都仅是补救作用，不可能将环境完全恢复到污染前的状态。因此，对待重金属污染最关键的应该是从源头着手，加强对相关污染企业的管理，从总量上控制重金属污染物的排放；同时应抓紧制订重金属污染综合整治方案，以更加积极有效地应对突发性重金属污染事件。

1.1.3.5 我国土壤污染状况专项研究

2006 年 7 月，我国投入了 10 亿元专项资金，开展全国土壤污染状况调查工作。这具有十分重要的现实意义，是制定土壤污染防治对策、做好土壤污染防治工作的基本前提。该专项研究历时 3 年半。其目的是通过分析土壤中重金属、农药残留、有机污染物等项目的含量及土壤理化性质，结合土地利用类型和土壤类型，摸清全国土壤环境状况，掌握土壤污染情况，研究基于土壤环境风险的土壤环境质量评价。截至 2010 年底，全国共采集土壤、农产品等各类样品 213754 个，获得有效调查数据 495 万个，点位环境信息数据 218 万个、照片 21 万张，制作图件近 11000 件，建成全国土壤污染状况调查数据库和样品库，数据总量达 1 TB，入库样品数量为 54407 份。该专项还组织完成全国土壤污染状况调查总报告和专题报告。针对重金属类、石油类、多氯联苯类、化工类污染场地和污灌区农田土壤等开展试点研究，完成 12 项试点工程、18 份研究报告和 7 部污染土壤修复技术指南草案，最后完成了《土壤保护战略研究报告》。

我国土壤污染状况调查工作还组织开展了土壤污染防治立法调研，起草了《土壤污染防治法》文本草案及法律条款编制说明；提出了中国土壤环境保护标准体系框架建议，形成了《土壤环境质量标准》修订草案。这是自《土壤环境质量标准》GB 15618—1995 颁布以来，二十多年后的第一次修订。土壤环境质量标准修订工作是一项艰巨而富有挑战性的任务。作为标准，其制定是一项十分严肃的工作，是在大量前期工作的基础上进行的，应具有科学性、政策性、实用性和可操作性。这不仅应有科学的制定方法，还应有大量资料，尤其是国内资料。

1.1.3.6 对国内相关研究的思考

国内对“水域-土壤-植被-大气”系统尺度上的重金属污染生态环境宏观研究较少，急需构建基于重金属元素循环迁移机理模型、遥感和 GIS 等先进技术的现代化立体监测系统与生态风险管理系统。这将是宏观尺度重金属污染监测和治理研究的重要发展方向。

对土壤重金属污染进行分析和评价，进而预测污染的发展趋势，制定合理的防污治污方法，这些研究对生态环境保护具有非常重要的意义，也是目前重金属污染研究的热

点之一。一方面,分析现状需要大量的基础数据;另一方面,评价需要统一的标准,而我国新的标准和评价方法尚未出台。在本书后文中,将不对术语"土壤环境质量评价"和"土壤重金属污染评价"做严格区分,可以理解成是对土壤环境质量的专项评价,即对特定的重金属污染物的评价。

对土壤重金属污染空间变异进行研究,有利于对植物生产进行合理布局,有效减少对食物链的污染。但是,由于土壤本身并非一个匀质体,而是一个具有高度空间连续性的变异体,因此,土壤类型的分布、土壤污染的演变都具有高度的空间异质性,也就是说,土壤组分的性质、污染程度的变化随着空间位置不同而发生较大变异。高度复杂的空间异质性,使得我们对土壤中重金属污染变化的空间动态的研究变得十分困难。这将是今后研究的重点和难点。

结合未来土壤学的新技术,应考虑到土壤中重金属污染物的迁移与城市空间形态特征密切相关,现代 GIS 技术能为土壤污染的空间形态研究提供有利的分析工具;土壤中重金属污染物的迁移和时-空演变是一种具有空间特征的复杂系统,而智能信息处理技术适用于模拟复杂系统和预测发展趋势。

1.1.4 GIS 技术

地理信息系统(Geographic Information System,简称 GIS)是随着地理科学、计算机技术、遥感技术和信息科学的发展而诞生并迅速发展起来的一门新兴技术。从计算机科学的角度而言,GIS 是一个能够对空间相关数据进行采集、管理、分析和可视化输出的计算机信息系统。

GIS 技术萌芽于 20 世纪 60 年代初,1972 年加拿大地理信息系统全面投入运行,这标志着第一个 GIS 系统的产生。此后,地理信息系统在全球范围内获得了快速发展。目前,在西方发达国家,GIS 应用已经渗透到社会经济生活的各个方面,包括资源管理、环境保护、灾害预测、投资评价、城市规划建设、人口和商业管理、交通运输、石油和天然气、教育、军事等众多领域。

随着国内经济建设的迅速发展,GIS 在城市规划管理、环境保护、交通运输、防灾减灾、农业、林业等领域发挥了重要作用。具体来说,GIS 技术在环保领域的应用包括以下几个方面:

1)应用 GIS 制作环境专题图。与传统周期长、更新慢的手工制图方式相比,利用 GIS 建立起地图数据库,可以达到一次投入、多次产出的效果。它不仅可以输出全要素地形图,而且可以根据用户需要分层输出各种专题图。GIS 的制图方法比传统的人工绘图方法要灵活得多,快速高效,并可以根据实际需要从符号和颜色库中选择图件,使之更好地突出专题效果和特性。

2)应用 GIS 建立各种环境地理信息系统。使用 GIS 技术建立各种环境空间数据库,把各种环境信息与其地理位置结合起来进行综合分析与管理,以实现空间数据的输入、查询、分析、输出和管理的可视化。

3)GIS 应用于环境监测。利用 GIS 技术可对实时采集的数据进行存储、处理、显示、

分析，直观显示和分析环境现状、污染源分布、环境质量评价，追踪污染物来源，为环境决策提供辅助手段。

4）GIS 应用于自然生态现状分析。利用 GIS 可以比较精确地计算生态量，客观地评价生态破坏程度和波及的范围。通过生态环境的空间分布与空间统计分析生态环境质量状况和生态环境变化的空间规律特点，为各级政府进行生态环境综合治理提供科学依据，为该地区经济的可持续发展与资源环境的可持续利用提供科学依据。

5）GIS 应用于环境应急预警预报。建立重大环境污染事故区域预警系统，对事故风险源的地理位置及其属性、事故敏感区域位置及其属性进行管理，提供污染事故的大气、河流污染扩散的模拟过程和应急方案。

6）GIS 应用于环境质量评价和环境影响评价。充分利用 GIS 的集成性，管理与生态环境密切相关的数据，利用 GIS 的空间分析功能，可以综合性地分析建设项目各种数据，帮助确立环境影响评价模型。由于 GIS 系统具有分层的结构，可将不同的环境影响进行计算并叠加，因此，GIS 是综合分析评价的有力工具。

环境信息系统在环境保护管理和决策工作中发挥着越来越重要的作用。而 GIS 技术的出现为环境保护工作迈向信息化、现代化提供了技术支持。GIS 技术在环境保护中应用的技术要点是建立环境 GIS 专用系统。其实质是在数据库管理系统（DBMS）和计算机辅助下，融合两个比较成熟的软件技术，并附加了对空间数据进行管理和分析的特殊功能。GIS 技术以其混合数据结构和独特的地理空间分析功能而别具一格。它还采用工业标准的组件对象模型（COM）技术，支持二次开发，从而为用户进行功能和结构定制拓展了空间。

综上，利用 GIS 结合智能信息处理技术来研究土壤重金属污染土壤质量评价指标和安全预警技术体系，能够在可视化环境对土壤重金属污染数据进行分析，准确地对污染情况做出评价，实现土壤重金属污染的预警预测。进而为深入研究土壤环境质量、合理利用土壤资源提供参考，对于政府制定针对性措施、平衡城市化发展和实施土壤环境保护具有非常重要的现实意义。

1.2 土壤重金属污染评价方法

土壤重金属污染评价是在对土壤现状及基础数据定量研究的基础上，对污染程度评价的定性研究，具有现实性和参考性。

1.2.1 土壤重金属污染分析工作流程

评价土壤重金属污染程度，首先应该对土壤中的各种污染物含量进行分析，这属于土壤环境评价监测工作的内容。这项工作通常以 GB 15618—1995《土壤环境质量标准》为原则，依据 HJ/T 166—2004《土壤环境监测技术规范》，由专业技术人员对常规项目、特定项目、选测项目的土壤实施监测。监测频次从“必要时”到“3 年一次”不等，但不应低于 5 年一次。

土壤污染物分析具体的工作流程如图 1-1 所示，主要包括野外调查和实验室分析两个阶段。

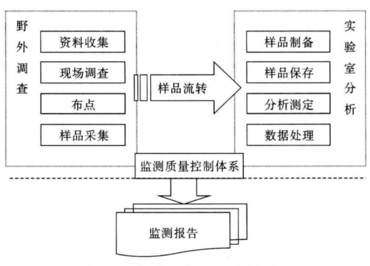


图 1-1　土壤污染物分析工作流程图

野外调查环节的工作重点是采集土样。出野外前要组织人员学习技术规范，并准备相关资料，包括交通图、土壤图、地形图等，还要有相关的土壤环境研究资料、历史资料、水文资料、气候资料等。现场调查可能需要多次踏勘，将调查得到的信息进行整理和利用，丰富采样工作图的内容。

野外工作需要携带的器具包括：①工具类，如铁锹、铁铲、取土钻（圆形、螺旋形）、刮刀、皮榔头及特定采样工具等。②器材类，如 GPS、罗盘、照相机、卷尺、样品袋、样品箱等。③文具类，如标签纸、采样记录表、铅笔、资料夹等。④安全防护用品，如工作服、安全帽、工作鞋、药品箱等。⑤交通类，如运输车辆。

野外现场的布点原则是：采样点的土壤类型特征明显，所采土样既可以作为采样区域总体的代表，土样间又互相存在一定程度的异质性。常用的布点方法有：简单随机布点法、分块随机布点法和系统随机布点法。样品采集过程可以采集表层土样，也可根据调查精度要求采集土壤剖面。采集过程注意严格按照《土壤环境监测技术规范》所要求的各环节规程，并注意填写样品标签和采样记录。

野外调查环节获得的土样经过样品流转环节，进入实验室分析。在流转过程中，工作人员需注意装运前核对，运输中防损，最后还要交接善后。

实验室分析的首要步骤是样品制备。这项工作主要通过制样、风干、研磨、过筛，获得可以用于仪器实验分析的土样。工作人员要按照样品名称、编号和粒径对土样分类保存。土样实验前的预处理方法有酸溶解法、碱熔法、高压密闭分解法、微波炉消解法、干灰化法和溶剂提取法。土壤中污染物的分析测定常用仪器有原子吸收光谱仪、分光光度仪、气相色谱仪、液相色谱仪等。分析测定后，要对实验数据进行必要的加工处理，由质控人员保证实验室测定数据的精密度。只有“平行双样测定”结果符合误差范围要求的

数据才能进入最终监测报告。质量控制体系在土壤污染物分析工作的整个过程都起作用,可以提高监测数据准确性,减少监测不确定性,保证监测误差可度量。

监测报告的撰写有规范格式,必须包括报告名称、实验室名称、报告编号、报告页标志、采样地点名称、采样时间、分析时间、检测方法、监测依据、评价标准、监测数据、单项评价、总体评价、监测仪器编号、检出限、采样点示意图、采样者、分析者、报告编制、复审者、签发者等信息。

1.2.2 土壤重金属污染评价方法

一种土壤环境质量评价方法,具有三个要素:评价因子、评价标准和评价模式。评价因子数量与项目类型取决于监测的目的和现实的经济和技术条件。评价标准常采用国家土壤环境质量标准、区域土壤背景值或部门(专业)土壤质量标准。评价模式常用污染指数法或者与其有关的评价方法。

下面介绍4种常用的土壤环境质量评价方法。

1.2.2.1 污染指数、超标率(倍数)评价

污染指数和超标率等计算公式如下:

$$\text{土壤单项污染指数}=\frac{\text{土壤污染物实测值}}{\text{土壤污染物质量标准}}$$

$$\text{土壤污染累积指数}=\frac{\text{土壤污染物实测值}}{\text{污染物背景值}}$$

$$\text{土壤污染物分担率}=\frac{\text{土壤某项污染指数}}{\text{各项污染指数之和}}\times 100\%$$

$$\text{土壤污染超标倍数}=\frac{(\text{土壤某污染物实测值}-\text{某污染物质量标准})}{\text{某污染物质量标准}}$$

对土壤环境质量评价,一般以单项污染指数为主,指数小污染轻,指数大污染重。

当区域内土壤环境质量作为一个整体与外区域进行比较或与历史资料进行比较时除用单项污染指数外,还常用综合污染指数。

土壤由于地区背景差异较大,用土壤污染累积指数更能反映土壤的人为污染程度。土壤污染物分担率可评价确定土壤的主要污染项目,污染物分担率由大到小排序,污染物主次也同此序。除此之外,土壤污染超标倍数、样本超标率等统计量也能反映土壤的环境状况。

1.2.2.2 内梅罗污染指数评价(内梅罗评价)

在实际工作中,内梅罗污染指数(Nemerow)可以用于水体、土壤等污染程度的评价。例如国家技术监督局于1994年颁布实施GB 14848—2007《地下水质量标准》,并在2007年修订,行文中推荐使用内梅罗指数法评价地下水的质量。

国家环境保护部于2004年推行的行业标准HJ/T 166—2004《土壤环境监测技术规范》也将内梅罗指数法列为土壤环境质量评价方法之一。

内梅罗污染指数 P_{Nemerow} 的计算公式为:

$$P_{\text{Nemerow}} = \sqrt{\frac{[\text{AVERAGE}(P_i)]^2 + [\text{MAX}(P_i)]^2}{2}}$$

式中 AVERAGE(P_i)表示单项污染指数的平均值,MAX(P_i)表示单项污染指数的最大值,P_i代表某种污染物的单项污染指数。

内梅罗指数反映了各污染物对土壤的作用,同时突出了高浓度污染物对土壤环境质量的影响。按照内梅罗指数评价土壤污染的标准见表 1-1。

表 1-1　土壤内梅罗污染指数评价标准

等级	内梅罗污染指数	污染等级
Ⅰ	$P_{\text{Nemerow}} \leqslant 0.7$	清洁(安全)
Ⅱ	$0.7 < P_{\text{Nemerow}} \leqslant 1.0$	尚清洁(警戒线)
Ⅲ	$1.0 < P_{\text{Nemerow}} \leqslant 2.0$	轻度污染
Ⅳ	$2.0 < P_{\text{Nemerow}} \leqslant 3.0$	中度污染
Ⅳ	$P_{\text{Nemerow}} > 3.0$	重污染

1.2.2.3　背景值及标准偏差评价

用区域土壤环境背景值(x)95% 置信度的范围($x \pm 2S$)来评价土壤环境质量的方法是:

若土壤某单项污染物 i 的监测值 $x_i < x - 2S$,则属于低背景土壤。

若土壤某单项污染物 i 的监测值在 $x \pm 2S$,则该单项污染物含量正常。

若土壤某单项污染物 i 的监测值 $x_i > x + 2S$,则土壤已受污染,或属于高背景土壤。

1.2.2.4　综合污染指数评价

综合污染指数(*CPI*)包含了土壤元素背景值、土壤分类标准尺度因素和价态效应综合影响。其表达式为:

$$CPI = X * (1 + RPE) + Y * DDMB/(Z * DDSB)$$

式中 *CPI* 为综合污染指数,X,Y 分别为测量值超过标准值和背景值的数目,*RPE* 为相对污染当量,*DDMB* 为元素测定浓度偏离背景值的程度,*DDSB* 为土壤标准偏离背景值的程度,Z 为用作标准元素的数目。主要有下列计算过程。

1)计算相对污染当量(*RPE*):

$$RPE = [\sum_{i=1}^{N} (c_i / c_{is})^{1/n}]/N$$

式中 N 是测定元素的数目,c_i是测定元素 i 的浓度,c_{is}是测定元素 i 的土壤标准值,n 是测定元素 i 的氧化数。对于变价元素,应考虑价态与毒性的关系。

2)计算元素测定浓度偏离背景值的程度(*DDMB*):

$$DDMB = [\sum_{i=1}^{N} (c_i / c_{ib})^{1/n}]/N$$

式中 c_{ib}是测定元素 i 的背景值。

3)计算土壤标准偏离背景值的程度(*DDSB*):

$$DDSB = [\sum_{i=1}^{Z}(c_{is}/c_{ib})^{1/n}]/Z$$

式中 Z 为用于评价元素的个数。

4)综合污染指数计算。

5)评价。用 *CPI* 评价土壤环境质量指标体系的标准见表 1-2。

表 1-2 综合污染指数(CPI)评价标准

X	*Y*	*CPI*	评价
0	0	0	背景状态
0	≥1	0<*CPI*<1	未污染状态,数值大小表示偏离背景值相对程度
≥1	≥1	≥1	污染状态,数值越大表示污染程度相对越严重

以上 4 种评价方法,在 HJ/T 166—2004《土壤环境监测技术规范》中有详细的说明。

对土壤重金属污染进行评价,依据土壤常规检测项目及分析方法的要求,主要考虑砷(As)、镉(Cd)、铬(Cr)、铜(Cu)、汞(Hg)、镍(Ni)、铅(Pb)、锌(Zn)8 种重金属元素。这 8 种重金属的背景值,可以参照 GB 15618—1995《土壤环境质量标准》中所给出的自然背景值,见表 1-3。

表 1-3 土壤环境质量标准值 (单位:μg/g)

项目	一级	二级			三级
	自然背景	pH≤6.5	6.5<pH≤7.5	pH>7.5	pH>6.5
砷(As) 旱地(≤)	15	40	30	25	40
砷(As) 水田(≤)	15	30	25	20	30
镉(Cd)(≤)	0.2	0.3	0.3	0.6	1.0
铬(Cr) 旱地(≤)	90	150	200	250	300
铬(Cr) 水田(≤)	90	250	300	350	400
铜(Cu) 农田等(≤)	35	50	100	100	400
铜(Cu) 果园(≤)	—	150	200	200	400
汞(Hg)(≤)	0.15	0.3	0.5	1.0	1.5
镍(Ni)(≤)	40	40	50	60	200
铅(Pb)(≤)	35	250	300	350	500
锌(Zn)(≤)	100	200	250	300	500

表中一级标准是为保护区域自然生态、维持自然背景的土壤环境质量的限制值；二级标准是为保障农业生产、维护人体健康的土壤限制值；三级标准是为保障农林业生产和植物正常生产的土壤临界值。

《土壤环境质量标准》中所给出的自然背景值和污染危害临界值有重要的参照价值。从实际操作和统一管理方面，制定标准意义重大，便于实施。但是，在实际应用过程中，考虑到我国地域广阔，土壤类型众多，土壤性质悬殊差异大，要制定一个全国范围内统一的土壤环境质量标准值，需要深入探究。从时间和发展趋势方面来说，这项标准是1995年制定的，远不能适应当前的要求。例如，污染物项目定得太少，尤其是铅的标准值偏低。已有专家着手对土壤环境质量标准进行修订，以求在对欧美等国的制定方法研究探讨的基础上，结合中国国情，建立更为健全的中国土壤环境质量标准体系。

1.2.3 对常见评价方法的分析

前述4种常见的土壤环境质量评价方法，各有优缺点。

污染指数、超标率（倍数）评价法和背景值及标准偏差评价法从统计学的角度出发，对监测样本进行了分析，得出污染程度的评价结果不具备评价区域间的横向对比。综合污染指数评价法的计算比较烦琐，重点考虑了重金属种类对污染的贡献。

内梅罗污染指数评价法应用最广，其优点是数学计算简洁，运算方便。对于某一个评价区域，先计算单项污染指数，再计算各种污染物的综合指数，根据分级标准，便可以评价区域的综合环境质量状况。但是，内梅罗污染指数评价法也存在着许多问题。例如，从计算公式中可以发现，评价方法过分突出了极大值对总体评价结果的影响，也就是说，待评价项目中即使只有一项指标值偏高，而其他指标值都较低，根据公式计算出来的内梅罗污染指数也会偏高。这就类似于"一票否决"的效果，显然不太客观。此外，在实际应用中，应该考虑不同评价因子对环境的毒性、降解难易及去除性难易程度等因素，不同污染因子的值应区别对待，也就是说要考虑不同单项污染指数的权重。

目前土壤重金属污染评价方法还有很多，除了《土壤环境监测技术规范》推荐的4种方法外，相关领域专家研究并提出的评价方法还有：几何均值综合评价法、污染负荷指数法、地积累指数法、沉积物富集系数法、潜在生态危害指数法、模糊数学法、灰色聚类法、基于GIS的地统计学评价法、健康风险评价法、环境风险指数法等。各种评价方法都有一定的适用范围，评价目的、优点、缺点不一而足。

总的来说，由于影响土壤环境质量的重金属种类及形态多样，加上地域差异等多种因素，各种因素的重要程度也不尽相同，用数学解析式来表示，存在一定的不足，表现在这些方法在评价过程中难以排除人员的主观性和评价系统的随机性，易造成评价结果失真和偏差。重金属评价因子与土壤质量等级间存在复杂的非线性关系，表现出极大的随机性、不确定性和非线性。

1.3 土壤样品的采集和制备

土壤样品的采集和制备大致分为两个阶段，第一个阶段是野外作业，第二个阶段是

实验室分析。根据《土壤环境监测技术规范》，进行土壤样品的采集与制备，具体工作步骤如下。

（1）采样前的准备

准备研究区域的地图（1：50000 或更小）、行政区划图、土壤类型分布图、重金属元素分布图、地球化学图、城市点源污染分布图、监测点位分布图等。

根据土壤监测野外调查工作环节的要求，需要携带的采样器具有木铲、刮刀、广泛 pH 值试纸、GPS、罗盘、照相机、卷尺、布制样品袋、样品箱、标签纸、采样记录表、铅笔、资料夹等。此外还要准备工作服、安全帽、工作鞋、药品箱等。如有需要，还应考虑野外踩点乘坐的交通工具。

（2）布点

经过现场踏勘后，设计布点方案。

布点方法很多，常用的为系统随机布点法。具体来说，就是将研究区域划分为网格，以 1 km^2的网格为一个采样布点基本单位，网格内采用四分法布点。选点的原则为在 1 km^2的范围内，尽可能选取具有相同使用功能的土壤类型。

（3）样品采集

土壤样品采集的具体方法是：先拨开覆盖在表层土壤最上层的树枝、碎石和杂草，用木铲垂直于地表，挖出 0～20 cm 的土壤剖面表层，取厚度大致相同的土壤薄片，观察土壤特征，除去大颗粒杂质，取 1 kg 左右，装入采样布袋中，扎口、编号。

如果研究土壤剖面情况，应该进行钻孔取样，人工打钻 2 m 深，根据需要分段采集不同深度的土壤样品，通常取距地表 20 cm、40 cm、60 cm、80 cm、100 cm、150 cm、200 cm，共 7 层。

（4）样品保存与流转

样品进行装样和编号时，必须同时写两个标签，装样袋内外各一个标签，注明采样地点、层次厚度、天气、时间、采集人等。

在实验室分析之前，应该对土壤样品进行初步加工。方法是：将野外采集来的土壤样品放在通风处，拣去根系及残枝落叶，摊开风干，避免暴晒。等待自然风干后，用塑料锤将土壤样品捣碎。如果是四分法采样，需要将一个采样网格内 4 个采样点采集的土壤样品混合，每个样品中取出大致相同质量的土壤样品，均匀混合，组合成一个混合样，重新编号。

最后，将混合样送到土壤实验测试中心进行分析测定。

（5）样品制备

重点分析测定土壤样品中的重金属含量时，在样品采样、风干、研磨、过筛、运输、贮存等环节都要特别注意，不能接触金属器具，以防土壤样品被污染。

在实验室中，将样品平铺在制样板上，用木棍或塑料棍碾压，并将植物残体、小石块等杂质剔除干净，细小已断的植物须根可采用静电吸附的方法清除。若在土壤中有铁锰结核、石灰结核、铁子或半风化体，不能用木棍碾碎，应细心拣出称量保存。

最后，将土壤混合样再次研磨，过孔径 74 μm（200 目）尼龙网筛，送检测室等待测试。

(6)实验室分析

经过土壤实验测试中心专业人员工作,分析测量土壤混合样品中的8种重金属元素含量。

为了保证数据的准确性,在样品测试过程中,可以将每10个样品以密码形式加国家一级标准物质(GSS土壤系列)进行质量监控,由此确保分析数据的准确可靠。

1.4 常见的微量元素分析仪器

在实验室中,进行土壤重金属元素分析,可以选择的微量元素分析仪器很多,按照仪器的工作原理可以分为原子荧光光谱类和等离子体质谱类。前者常见的仪器型号有AFS-230和XGY-1011,后者常见的型号有ICP-MS和Excell-TJA。

AFS-230是一种原子荧光分光光度计。该仪器采用氢化物发生原子荧光光谱吸收测定,具有原子吸收和原子发射光谱两种技术的优势,分析灵敏度高、干扰少、线性范围宽,能够实现多元素同时分析。可以用于测定土壤样品中的砷(As)、锑(Sb)、铋(Bi)、汞(Hg)、硒(Se)等微量元素的含量。

XGY-1011是一种原子荧光光度计。该仪器为单道仪器,采用氩氢火焰低温自动点燃装置及分析技术,原子化器无需屏蔽即可提高灵敏度和稳定性,操作安全。此外,采用氢化物/汞蒸气及气液分离技术,有效地避免了液相干扰问题,数据的稳定性和可靠性高。

ICP-MS和Excell-TJA是由美国TJA公司生产的电感耦合等离子体质谱仪。该仪器的工作原理是:待测物质以气溶胶或气体形式进入高频电场,在快速变化的电场作用下形成离子,通过取样锥和截取锥内负气压将所形成的离子吸入真空室,在变化的电场作用下,将各种离子按其质荷比 m/Z 分离开,最后由检测器计数结果来计算出被测项目的浓度。可以利用该仪器分析土壤样品中的微量元素,也可以在使用分离技术后,利用该仪器测定土壤中重金属元素的形态。

第2章

GIS支持的土壤重金属污染评价

研究土壤重金属污染评价新方法、新技术,能为制定环境保护政策提供依据,是合理利用资源、保护城市生态环境的前提。为了实现一种新的土壤重金属污染评价支撑平台,本章详细介绍GIS应用开发的技术。借助GIS技术开发环境污染评价系统,该平台还可以对大气、植物、水体等环境要素实现监测,为城市环境综合质量监控提供基础数据,并着重实现对土壤重金属污染的分析和评价。本书以介绍土壤重金属污染研究和评价为主,对大气、水体、植物的重金属研究有所涉及并简要介绍。

2.1 土壤重金属污染评价系统

综合运用GIS与人工智能、数据挖掘等技术是实现土壤重金属污染评价的新途径。本节主要介绍GIS支持的环境污染研究平台基本框架及功能。

2.1.1 基本框架

在充分学习国内外研究现状的基础上,通过对重金属污染监测及预测关键技术深入细致的研究,提出土壤重金属污染评价及安全预警系统的基本框架,如图2-1所示。

大体上来说,该系统的底层数据库有3个,分别是空间数据库、属性数据库和模型数据库。通过对基础地理要素的预处理,建立空间数据库,这是实现空间分析服务的基础数据。属性数据库支持多源异构数据的集成,经过采集和预处理监测数据,系统地将城市环境的多要素数据规范地存储到数据库中。考虑到模型的重用性和扩展性,建立模型数据库,通过系统调用和优化,为评价和预测预警提供研究数据。

从功能上来说,该系统有两个模块,一个是土壤重金属污染现状评价模块,并呈现污染程度空间分析的结果;一个是安全预警和决策支持模块,通过空间分析技术为污染源判别和灾难链阻断。

该系统的技术实现方案可以概括为:空间数据预处理,GIS平台做支撑,属性数据多源化,评价、预测靠模型。充分利用GIS技术的集成性来开发土壤重金属污染土壤质量评价和安全预警系统,一方面可以充分挖掘空间数据和属性数据,借助多种污染评价模型为土壤重金属污染数据的现状分析提供可视化研究平台,为准确地对污染情况做出评价提供依据。另一方面,在空间因子分析、空间相关分析、空间变异分析的辅助下,借助

模型实现土壤重金属污染的预警预测，对于污染源判别和灾难链阻断的决策有参考和支持作用。

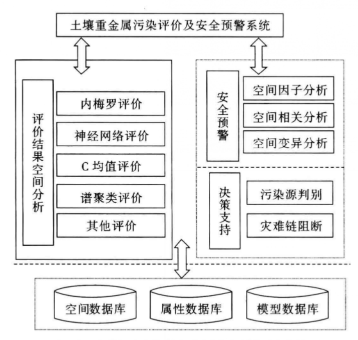


图 2-1　土壤重金属污染评价及安全预警系统框架

2.1.2　主要功能

以土壤重金属污染评价及安全预警系统为基础平台，以监测数据及评价结果为依据，可以为深入研究土壤环境质量、合理利用土壤资源提供参考，这对于政府制定针对性措施、平衡城市化发展和实施土壤环境保护具有非常重要的现实意义。

土壤重金属污染评价及安全预警系统的主要功能，如图 2-2 所示。

文件操作功能，主要是对工作空间的管理。工作空间是由若干空间数据的图层组成的，在此之上可以关联属性数据，如大气采样信息、植物采样信息、土壤采样点信息、土壤理化参数、土壤中重金属含量数据等。创建工作空间可以集成多种数据，保存工作空间可以方便再次打开、使用、退出。

视图操作功能，主要是对地图的操作。可以像 Google 地图中的使用一样，对地图进行放大、缩小、全图、平移、鹰眼等操作，还可以控制图层的显示或隐藏。

统计分析功能，除了汇总基本的统计量，如最大值、最小值、平均值、中值、标准差、方差等，还可以进行多变量的相关分析和因子分析，为深入研究土壤中多种重金属提供了数据分析基础。对统计得到的结果，还可以以柱状图、饼图、百分比堆积柱状图等方式呈现。

GIS 分析功能，是在空间数据的支持下完成对污染情况的空间分析。包括按照给定条件查询空间要素，查询框选区域内的要素，量算两个地理要素间的距离，对空间数据进

行 Kriging 插值运算,绘制专题图等。系统实现的专题图有多种,如区域单一重金属的含量专题图、区域污染评价等级专题图、预测区域若干年后污染情况等值线专题图。

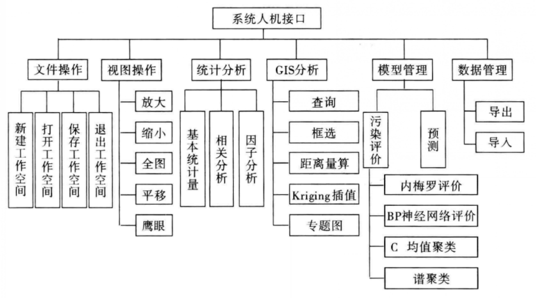


图 2-2　系统主要功能

模型管理功能,是对系统中用到的所有模型进行统一管理和调用。主要模型有两类:污染评价模型和预测模型。评价模型是重点,有内梅罗评价模型、BP 神经网络评价模型、C 均值聚类模型、谱聚类模型等。评价模型可以对单一采样点进行评价,也可以在多点评价后,通过插值实现对区域污染情况的评价。

数据管理功能,完成土壤重金属污染监测数据、空间图层数据、模型、评价结果等的导入及导出。对数据实施规范统一的管理,形成属性、空间、模型三大数据库,是实现系统各种功能的基础和核心,为 GIS 空间分析和评价,对土壤重金属进行现状分析和风险评价,查询、统计分析和制图制表等功能提供数据基础。通过数据管理,可以建立准确、一致的信息来源,为揭示土壤中重金属的分布特征、污染成因及空间变异规律,实现土壤重金属监测数据管理的科学化提供必要支持。

2.1.3　数据库设计

建立起支持多源异构的数据库,才能够对某区域中土壤重金属污染情况进行多样化的空间分析,这是快速准确地找到污染源并实现土壤重金属污染预警预测的基础。

土壤重金属污染评价及安全预警系统的基础底层数据是多源的,如图 2-3 所示。

从数据的多源来看,主要有两大类,分别是空间数据和属性数据。空间数据最大的特点是含有地理坐标信息,也就是我们平常所说的矢量数据,表现为点、线、面要素,通常以图层的方式组织,需要专用的空间数据库来表达地理要素直接的拓扑关系,以进行空

间运算，如交、割、叠加等。属性数据从本质上来说是文本数据，包括数字、文字、图片等，可以进行一定的运算，通常存储在属性数据库中。在土壤重金属污染评价及安全预警系统中，属性数据最大的特点是需要与空间数据关联起来，作为空间分析的运算对象。

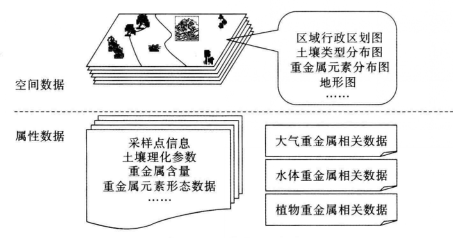


图 2-3　多源数据示意图

系统的空间数据库主要包括矢量电子地图，如行政区划图、土壤类型分布图、重金属元素分布图、地形图等多源数据。属性数据库主要包括：采样点信息、土壤理化参数、土壤中所含主要重金属含量、重金属的形态数据等，此外还有大气、水体、植物中的重金属相关数据。

从数据的处理和流转上来说，可能处理和使用系统数据的人员也是多样化的，系统的顶层数据流如图 2-4 所示。从图 2-4 可以看出，数据使用者主要有 5 类：样品采集员、样品分析员、模型分析员、科学工作者和系统管理员。

样品采集员主要负责土壤地球化学样品采集子系统的样品信息收集和整理，同时加工处理监测站和采样点信息，其中一个监测站包括一个或者多个采样点。

样品分析员主要负责土壤地球化学样品分析子系统中样品分析结果（包括理化参数、仪器分析、试验结果等）数据的加工。

模型分析员主要负责决策模型子系统中模型信息的收集和管理。

科学工作者主要为动态监测及预警预报子系统提供监测和决策信息，通过综合使用采集子系统、分析子系统和模型子系统，在时空信息分析的基础上给出客观性、时效性较强的土壤质量评价动态监测结果和污染源、灾害链有效阻断建设性决策方案。

系统管理员负责系统管理子系统的维护，确保整个系统的正常运行和更新发展。

对于不同的数据使用者而言，系统界面的友好和数据标准规范很重要。因此，应该分别设计不同的数据工作模块，力求使用户用起来方便、简单。此外，还要将数据存储文件的格式统一规范，这是最基础的工作。例如，采样点信息就应该包括这样一些数据项：采样点坐标位置、采样点名称、采样点经度、采样点纬度、采样点海拔、采样点覆盖范围、采样点备注信息等；土壤样品的信息就应该包括这样一些数据项：土样 ID、采样点 ID、采

样及分析过程 ID、土样剖面位、土样类型、土样利用类型、土样有机质含量、土样黏粒度、土样 pH 值、土样各种重金属元素含量、土样备注信息等。

如何借助 GIS 技术构建基础数据库，整合城市环境四大要素的多源数据，成为关键问题。总的来说，要设计好系统的数据库，必须坚持"面向用户，规范易用"的原则，重点解决在 GIS 支持下，整合多源数据，建立城市环境空间数据库和污染监测属性数据库，涵盖大气、水、土壤、植物的重金属元素含量监测信息。建立空间数据库，管理区域的行政区划矢量数据、重要地理要素分布矢量数据、采样点分布等矢量数据。建立属性数据库，管理土壤、植物、水、大气等各环境要素中重金属元素含量数据、重金属形态数据、重金属理化参数数据等。

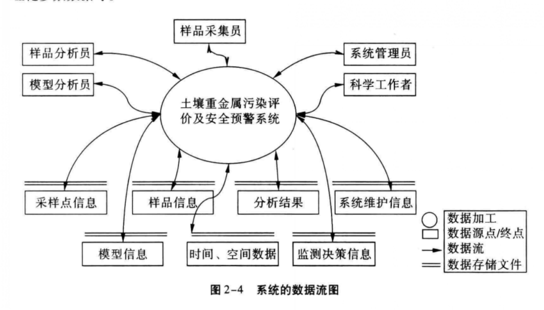


图 2-4 系统的数据流图

2.2 GIS 二次开发的准备

目前国内兴起的 GIS 应用开发大多数都是二次开发。对于实际应用来讲，二次开发的好处是显而易见的。表现在以下几个方面。

一是方便性。二次开发的技术含量较低，主要关注的不是 GIS 本身的技术，而是特定的用户需要的具体的应用，也就是我们常说的业务。

二是高效性。一般来说，开发 GIS 底层平台相对难度要大得多，而针对某种应用的 GIS 二次开发，充分利用现有的 GIS 平台实现具体的业务。这样就能大大减少工时，压缩开发成本，在时效和产出上保证效率。

三是稳定性。我们所进行的二次开发，是在现在流行的成熟的 GIS 平台之上。这些平台在产品化、市场化的过程中慢慢变得稳定、可靠。

但二次开发也不是十全十美，毕竟使用别人的 GIS 平台，就要为初期购置、后期升级投入资金，并且在开发中无法深入底层，只能在提供的接口之上搭建自己的系统，当需要

扩展时会变得束手无策。

本节重点介绍利用 GIS 产品为实现专门的实际应用，实现二次开发的主要环节和基本方法。

2.2.1 开发选型

随着地理信息互操作服务技术的发展，GIS 技术已经从初始的处理地理信息转换成为实现互操作和地理信息服务。目前有大量成熟的 GIS 软件平台，如 ESRI（Environment System Research Institute，美国环境系统研究所）的 ArcGIS，MapInfo 公司的 MapGIS，Autodesk 公司的 MapGuide，国内主要有超图公司的 SuperMap，中地数码集团的 MapGIS 等产品。这些产品大都支持二次开发，如何利用这些产品开发满足自己需要的 GIS 应用系统，已成为众多科研机构与企业共同关心的问题。

GIS 应用开发模式，常见的有三种，独立开发、宿主式开发、组件式开发。

独立开发模式，不依赖于任何 GIS 工具软件，从空间数据的采集、编辑到数据的处理分析及结果输出，所有的算法都由开发者独立设计，选用的程序设计语言有 Visual Basic、Visual C++等。该模式的开发周期长，软件功能相对简单，对开发者要求高，但用于购买软件的费用相对节省。

宿主式开发模式，基于某一 GIS 平台软件，进行应用系统开发，如 ArcView 中的 Avenue 语言，MapInfo Professional 的 MapBasic 语言。这种开发方式较为容易，但是二次开发的脚本语言功能较弱，所开发的系统不能脱离 GIS 平台，是由 GIS 软件平台解释执行的，运行效率低，用户界面受平台软件的限制。

组件式开发模式最为常见，其本质就是利用建立在组件技术基础上的 GIS 功能组件，实现 GIS 应用的各种功能。它的优点就是开发人员可以基于通用各种可视化开发工具，使用任意的编程语言（如 VB、VC、Delphi），实现基于电子地图操作的地理信息交互功能。组件通常都小巧灵活、价格便宜。

组件式 GIS 的基本思想是把 GIS 的各大功能模块划分为几个控件，每个控件完成不同的功能。各个 GIS 控件之间，以及 GIS 控件与其他非 GIS 控件之间，可以方便地通过可视化的软件开发工具集成起来，形成最终的 GIS 应用。控件如同一堆各式各样的积木，他们分别实现不同的功能（包括 GIS 和非 GIS 功能），根据需要把实现各种功能的“积木”搭建起来，就构成应用系统。

从计算机技术的角度来说，组件式 GIS 是面向对象技术和组件式软件在 GIS 软件开发中的应用。组件式对象模型（Component Object Mode）简称为 COM，是 OLE（Object Linking & Embedding）和 ActiveX 共同的基础。COM 不是一种面向对象的语言，而是一种二进制标准。COM 所建立的是一个软件模块与另一个软件模块之间的链接，当这种链接建立之后，模块之间就可以通过称之为“接口”的机制来进行通信。COM 标准增加了保障系统和组件完整的安全机制，扩展到分布式环境，则被称作 DCOM（Distribute COM），可以实现 COM 对象与远程计算机上的另一个对象之间直接进行交互。ActiveX 是一套基于 COM 的可以使软件组件在网络环境中进行互操作而不管该组件是用何种语言创建的

技术。ActiveX 控件是 ActiveX 技术的重要内容，是一种可编程、可重用的基于 COM 的对象。ActiveX 控件通过属性、事件、方法等接口与应用程序进行交互。

几个著名的 GIS 软件公司把 COM 技术应用于 GIS 开发，纷纷推出由一系列 ActiveX 控件组成的组件式产品。有代表性的 GIS 开发组件有：ESRI（Environment System Research Institute，美国环境系统研究所）的 MapObjects，MapInfo 的 MapX，ArcGIS 的 ArcObjects，超图公司的 SuperMap Object 控件，中地数码集团的 MapGIS ActiveX 等。

下面仅对历史最为悠久、国内外市场占有率较大的 MapObjects 和 MapX 进行对比，见表 2-1。

表 2-1　MapObjects 和 MapX 的对比

功能	MapObjects	MapX
显示的地图数据格式	ArcView 的 SHP、ARC/INFO 的 Coverage、SDE 图层	MapInfo 的数据格式
叠加栅格图像	支持	支持
对地图的常用操作	放大、缩小、漫游等	放大、缩小、漫游等
图层控制	增加、移除、设置当前层	增加、移除、设置当前层
属性数据绑定	支持	支持
地图信息查询方式	1. 通过鼠标选取特征； 2. 通过 SQL 查找特征； 3. 通过空间操作选取要素	1. 通过鼠标选取特征； 2. 通过 SQL 查找特征； 3. 通过空间操作选取要素
专题图	较弱	支持
GPS 集成	支持	支持
用户绘图图层	无	支持
生成/编辑地图对象	较弱	较弱
地图标注	支持	支持
地图符号化	较弱	较弱
分析功能	支持	无
地理编码	支持	支持
开发语言	VC、VB、PowerBuilder、Delphi 等	VC、VB、PowerBuilder、Delphi、LotusNotes 等
版本和价格	最新版本是 MapObjects 2.4。对合作伙伴和最终用户有不同的折扣，价格需咨询 Esri 中国（北京）有限公司	最新版本是 MapX 5.0。开发包与客户端 License 搭配销售，价格需咨询 PBBI 公司大中华区总部（Mapinfo 已被 Pitney Bowes 公司收购）
开发文档	齐全	齐全

实施 GIS 二次开发都要有一个选型过程。要在众多的 GIS 软件平台中，选出适合特定开发环境的某个公司的组件产品作为开发基础，这项工作有一定的难度。在选型时，需要重点考虑以下几个方面：

首先，必须能够满足项目的实际需求，平台支持不能弱，功能点不能少。但是也不必一味求大求全，因为出于对资金投入的考虑，没有必要为总也用不到的功能买账。

其次，在满足功能要求的情况下，要从长远看，能够保证此项投资的可持续发展性，即买来的东西不能仅限于当时好用，以后的可扩展性、可升级性也需要考虑。

最后，要考虑性价比。价格合适，并且开发文档齐全。

综合以上几个方面的要素，MapObjects 可以满足土壤重金属污染评价研究所需要的 GIS 功能，可以确定为二次开发的 GIS 地图功能的实现基础。在以后的内容中，会比较详细地解释 MapObjects 这一控件。

至于开发平台和程序设计语言的选择，目前比较流行的两大开发阵营，分别是 Java 和.NET。因此，不论是选这两大主流哪一个，都可以作为 GIS 二次开发的基础平台。推出 MO 产品的 ESRI 开发公司一直都很注重产品对 Java 和.NET 两大开发平台的支持，提供有相应的开发包及文档。在实际应用中，开发者可完全根据自身的技术专长和应用能力，选择一种便于上手的开发平台和程序设计语言。

以微软公司的.NET 平台来说，由于.NET 框架是一种新的计算平台，支持多种不同的编程语言，它以面向对象的思想简化了应用程序的开发，使得开发人员可以把精力集中在实现业务逻辑的核心上，无需过多关注程序设计语言及语言运行环境。此外，.NET 平台支持控件管理方式，具体来说，无论控件以何种形式存在，以何种方式调用，都可以无缝耦合到应用程序开发的过程中。也就是说，.NET 程序集是各种控件的载体，其本质是使用.NET 的反射机制来对控件进行管理和调用，这就为用户开发集成多种空间的应用程序，提供了极大的便利。

Visual Basic .NET 语言是一门语法简单、类型安全、灵活兼容的新型面向对象的编程语言。其语法的简单和易于理解使得开发人员能够非常轻松地使用，并且该语言具有灵活、高效的特性，是.NET 平台下的一种主流开发语言。

2.2.2 MO 概述

MapObjects，简称 MO 是全球最大的 GIS 软件供应商 ESRI 在业界最早推出的 GIS 软件组件，是当今流行的组件式 GIS 开发软件之一。目前最新的 MO 版本是 2.4，它是一组基于 COM 技术的地图应用开发组件，由一个称为 Map 的 ActiveX 控件（OCX）和近 50 个自动化对象组成，在标准的 Windows 编程环境下，能够与其他图形、多媒体、数据库开发技术组成完全独立的综合性应用软件，为基于 C/S 模式的应用业务提供良好的地图开发环境。

MO 的稳定性是其能被成功应用的重要前提和保证，它与 ESRI 的旗舰产品 ArcGIS 完全兼容。从结构简练的桌面数据格式（Shape 文件）到以拓扑关系为基础的经典 ArcInfo Coverage，以至基于数据库（DBMS）和 ArcSDE 基础之上的 Geodatabase，都能够被

MapObjects 兼容。除了矢量数据以外，MO 还能够读取多种格式的栅格数据，如 BMP，TIF，JPEG，ArcInfo Grid，Erdas Image 等。MO 本身也能够生成 Shape 文件格式的 GIS 数据，该格式已经成为 GIS 业界事实上的基于桌面应用的标准。采用 MapObjects，能够最大限度地与主流 GIS 技术融合，保护用户在数据生产、功能开发以及人员培训上所作的一切投资。

MO 以最少的接口提供了常用的 GIS 功能，甚至 GPS 的动态特性，同时做到了结构合理，简单明了，容易理解和扩展。与之相比，其他的 GIS 组件，不是由于提供的功能过多而破坏了其内在的结构美并影响了用户扩展的灵活性，就是结构过于简单而使功能大大弱化。可以说，MapObjects 在功能和结构两者之间，得到了完美平衡，体现出了软件的艺术性，能给人带来愉悦的感受。

从数据结构上来说，MO 对地图数据的组织方式概念清晰，易于理解。他将一个综合性的地图分解成多个图层，每个图层表达一类地理要素。图层数据来源广泛，既可以是 GIS 矢量图层，也可以是 CAD 图层，甚至影像数据。在数据文件内部，对于 GIS 和 CAD 的矢量图层，都统一用记录集（Recordset）来表达。记录集就像关系数据库中的二维表，行表示每个要素、列表示每个属性。如果含有名称为"Shape"的列，则该记录集表示的是地图要素，否则表示的是普通的数据库表格。这样，在数据结构内部就消除了 GIS 和 MIS 数据之间的差异，简化了程序员的数据观点。Shape 字段以面向对象的技术封装了要素的图形部分。它把图形划分为 3 种类型，即点、线、面。这些几何类型的组织，实际上是引入了部件（Part）的概念，复合两个以上的点构成一个部件，线和面再由部件构成。对于地图要素的注记，则被设置为属性字段的自动标注。标注的位置由 Shape 字段中的图形要素决定，它可以是点、线或面，通常是线。标注的内容由普通属性表示，既可以是字符型，也可以是整型或浮点型，通常是字符型。这样，一切要素既可以按图形表示，也可以按其属性内容标注显示，甚至可以两者同时显示，地图显示和制图的方式是极为灵活的。以 MO 2.2 为例，所包含的基本对象如图 2-5 所示。

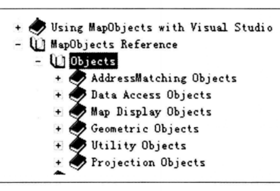


图 2-5 MO 的基本对象

这些基本对象分别是：地理匹配对象（AMO）；数据访问对象（DAO）；地图显示对象（MDO）；几何图形对象（GO）；实用对象（UO）；投影对象（PO）。

从 MO 对象所提供的功能的接口来看，常用的地图应用基本上都能实现，主要有：创建新的 Shape 文件；更新属性或图形数据；绘制点、线、椭圆、矩形和多边形等图形要素，显示多图层地图；绘制描述性的文本注记；地图上通过线、方框、区域、多边形、圆来提取图形要素；多种查询方式，如简单点查询、空间查询、相对位置查询、SQL 逻辑条件查询等；空间统计；地图缩放和漫游；丰富的绘图方式，如按值润色、分类显示、绘制密度图、产生各类专题图等；属性自动标注；显示航空和卫星遥感影像；动态显示实时或系列时间组数据，如 GPS 动态数据；地址匹配；投影变换；Buffer，Union，Intersect 等空间分析算子。

通过调用这些接口，用户能够开发简单的电子地图应用，甚至可以开发复杂的基于 GIS/GPS/RS 的 3S 应用。利用空间分析算子，可以构造出具有一定复杂度的空间分析模型。由于提供了数据更新接口，用户能够扩展出自己的编辑工具，使得二次开发能够满足更为广泛的应用要求。

2.2.3 开发环境

实施 GIS 二次开发的首要步骤是准备开发环境。土壤重金属污染评价及安全预警系统的开发选用 Visual Studio .NET 2003 作为开发平台，使用 MapObjects 组件，调用 Matlab、Surfer、Excel、TextNote 等软件进行综合的二次开发，实现了土壤重金属污染的分析评价及预测预警。表 2-2 详细列出了开发环境的硬件、软件清单。

由于开发环境中需要的软件比较多，尤其是在.NET 环境下安装 MO 需要特别注意，在这里详细说明一下准备开发环境的步骤，共有 5 个步骤。

(1)安装.NET 框架及 IDE 环境

运行“……\VSDOTNET”中的安装文件，来安装 Visual Studio .NET 2003。这里安装程序是镜像文件，因此，安装过程中有 3 个步骤。

表 2-2 系统开发环境

硬件环境(最低)	系统环境	软件环境
处理器：Pentium M 1.73 GHz	操作系统：Microsoft Windows XP Professional 2002 Service Pack 2	Microsoft Office Professional Edition 2003 (安装时需包括 Word、Excel)
硬盘：20 GB	软件开发 IDE：Visual Studio .NET 2003	MapObject 2.2 (安装时需包括 SDK For .NET)
DVD 光驱		Matlab R2007 (只需安装运行环境)
		Surfer 7
		Textnote

1)安装虚拟光驱“daemon341. exe”。

2)使用虚拟光驱映射到“sc_vs.NET_2003_enar_cd1. iso”，提示安装 Visual Studio .NET 2003，这时弹出对话框提示插入光盘安装预备环境。

3)使用虚拟光驱映射到“sc_vs.NET_2003_prereq. iso”，安装.NET 框架等内容。

(2)安装 Office Professional Edition 2003

运行“……\OfficeXp-Setup\OFFICE11”中的“SETUP. EXE”执行 Office Professional Edition 2003 的安装。注意选择 Word、Excel 为必选部分。

(3)安装 MapObject 2.2

运行“……\Mapobject_2.2”中的 Mo22Setup. EXE，执行 MO 的安装。需要特别注意的是，要在.NET 环境下正确安装 MapObject2.2 版本应遵循以下步骤。

1)修改注册表。单击开始菜单中的运行，输入 regedit，可以打开注册表编辑器。在注册表编辑器中，找到“HKET _ LOCAL _ MACHINE \ SOFTWARE \ Microsoft \.NETFramework”这个位置，如图 2-6 所示。然后单击鼠标右键，选择“新建字符串值”，名称设为“sdkInstallRoot”，(不要修改已有的“sdkInstallRootv1.1”)双击该字符串值，打开字符串编辑对话框，如图 2-7 所示，把其值设置到.NET 的安装路径下，如“C:\Program Files\Microsoft Visual Studio .NET 2003\SDK\v1.1\”。

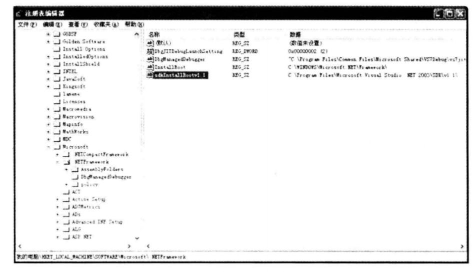


图 2-6　在注册表编辑器中找到修改配置

编辑字符串

数值名称(N):

sdkInstallRoot

数值数据(V):

\Program Files\Microsoft Visual Studio .NET 2003\SDK\v1.1\

确定　取消

图 2-7　新建字符串值

2)修改完注册表后，再安装 MapObject2.2，双击“MO2.2Setup.EXE”执行安装。注意，安装时“Select Components”，这一步必须将 MapObjects SDK for .NET 复选框上打勾选中，如图 2-8 所示。

3)安装后，需要注册 5 个与图层显示相关的 MO 组件。方法是用鼠标选中 COMCTL32.OCX，COMDLG32.OCX，grid32.OCX，MoLegend2DOTNET.OCX，tabct32.ocx 这 5 个文件，拖动到“RegSvr32.exe”中，即可完成注册。这 5 个组件在“……\ocx”文件夹下都可以找到，如图 2-9 所示。

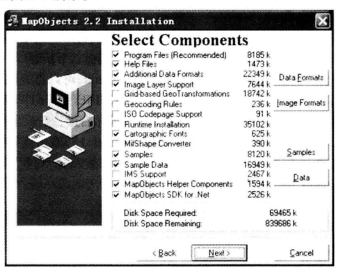


图 2-8　安装 MO 2.2

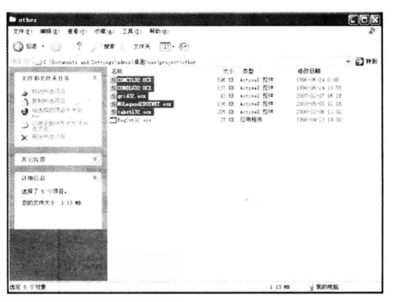


图 2-9　注册组件

4)正确安装 MO 后,可以再把注册表中的 sdkInstallRoot 删掉。

(4)安装 Matlab R2007a

使用 Matlab R2007a 开发 COM 组件 DrawSoilGraph_pkg,实现数据的曲线拟合。在没有安装 Matlab R2007a 的机器中需要安装 Matlab 的运行环境。运行"……\matlab"中文件来安装 Matlab 及其组件。共有 3 个步骤。

1)运行"MCRInstaller. exe"执行 Matlab R2007a 运行环境的安装。

2)双击"vcredist_x64. exe"或"vcredist_x86. exe"执行安装,不同的计算机可能不一样,只要有一个安装成功即可。

3)双击"……\matlab\DrawSoilGraph\distrib"文件夹中的"DrawSoilGraph_pkg. exe",3 次按 y 键并回车,如图 2-10 所示,直至出现的 DOS 界面关闭表示已经正确注册了组件。



图 2-10 安装 Matlab R2007a 运行环境

(5)安装 Surfer 7.0

运行"……\Surfer7"中文件来安装 Surfer7.0。双击打开"Surfer70233. exe",该软件的序列号存放在"serial. txt"中。最后一步选择"以后注册"完成安装,如图 2-11 所示。

(6)最后,重新启动电脑,完成开发环境的安装

在.NET 框架下,进行应用系统的开发,可以选择某种开发者熟悉的编程语言。下面以 Visual Basic .NET 语言为例,介绍新建一个名为"test"的项目,基本步骤为:

首先,启动 Visual Studio .NET 工作环境,方法是单击"开始菜单→程序→Microsoft Visual Studio.NET 2003→ Microsoft Visual Studio.NET 2003",进入编程开发环境后的界面,如图 2-12 所示。

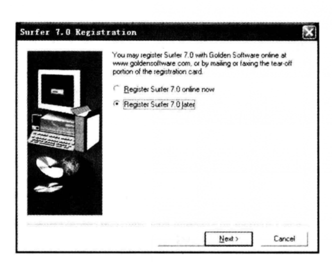


图 2-11　安装 surfer 7.0

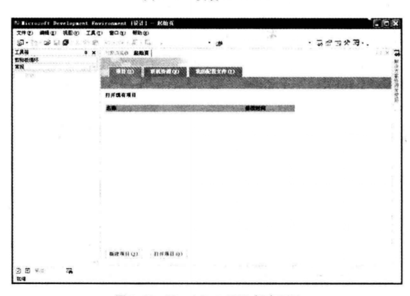


图 2-12　Visual Basic.NET 新建项目

单击"文件——新建——项目"后，弹出"新建项目"对话框，如图 2-13 所示，在项目类型中选择"Visual Basic 项目"，在右侧模板选择"Windows 应用程序"，在"名称"后的文本框中填写项目的名称"test"，在"名称"后的下拉菜单中选择项目所在的位置。

选择完毕后，点击确定，就完成了新建一个 Visual Basic 项目的工作。打开了一个项目的窗体设计区域，如图 2-14 所示。

新建了项目之后，要利用 MO、Surfer 等进行二次开发，首先要正确引用 MO 和 Surfer，其方法步骤如下。

(1)在项目中正确引用 MO 2.2

1)打开项目文件夹"……\project\test"中的"test. sln"文件，在.NET 中"项目"中选择"添加引用"子菜单，如图 2-15 所示。

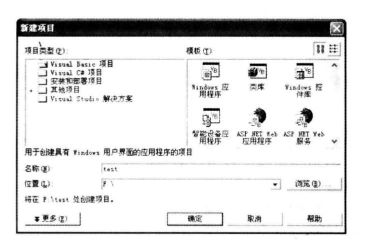


图 2-13　新建项目的基本信息

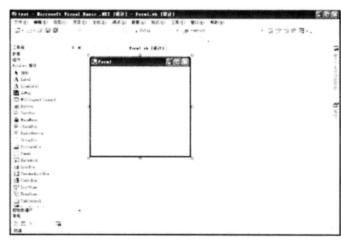


图 2-14　窗体设计区域

图 2-15　在.NET 中“项目”中选择“添加引用”子菜单

2）选择“添加引用”对话框中的“项目”选项卡，单击“浏览”按钮，指向系统中安装好的 MO 2.2，如图 2-16 所示，完成系统对 MO 2.2 的引用。

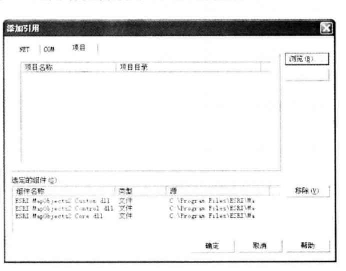


图 2-16　在“项目”选项卡中引用 MO 2.2

3）检查是否正确引用了 MO 2.2，可以查看“.NET”选项卡。如果正确引用，会出现如图 2-17 所示窗口。

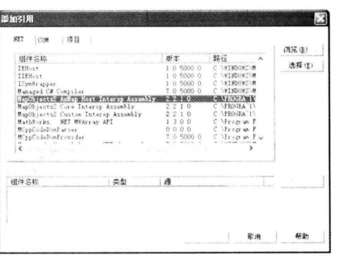


图 2-17　“COM”选项卡中查看正确引用了 MO 2.2

（2）在项目中正确引用 Surfer 7.0

1）打开项目文件夹“……\project\test”中的“test. sln”文件，在.NET 中“项目”中选择“添加引用”子菜单。选择“添加引用”对话框中的“项目”选项卡，单击“浏览”按钮，指向系统中安装好的“Surfer. exe”，如图 2-18 所示，完成系统对 Surfer 的引用。

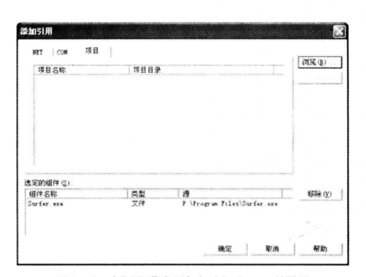


图 2-18　在“项目”选项卡中对 Surfer.exe 的引用

2）检查是否正确引用了 Surfer，可以查看“COM”选项卡，如果已经正确引用，会出现如图 2-19 所示的窗口。

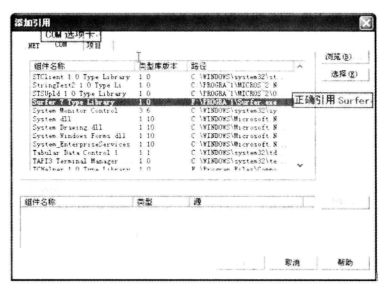


图 2-19　“COM”选项卡中查看正确引用 Surfer

至此，就配置好了项目的开发环境，可以进行代码编写了。

第3章 GIS基本功能的实现

GIS二次开发最基础的是电子地图功能的实现,包括电子地图的显示、操作、处理等。这些功能的实现离不开电子地图数据,也就是图层数据文件。本章介绍如何实现GIS二次开发中地图交互操作的开发。

3.1 地图功能

产生图层数据文件的方法很多,可以从ArcInfo中产生,也可以从其他GIS软件,如MapInfo、SuperMap、MapGIS等系统中转换而来。在MO中可以使用的数据文件格式主要有Shapefile、Coverage、SDE、VFP和各种CAD文件。

Shapefile文件采用早期的dBase数据库管理模式,一个文件夹是一个数据库,文件夹的不同文件就是数据库中的表,每个表代表GIS图中的一个图层。每个图层文件必须包括3个以上的特征文件才能在MO中使用,这3个特征文件分别是:空间特征数据文件(即图形文件.shp),索引文件(.shx),属性数据文件(.dbf)。

每一个shp文件包含一种类型,如点、线、多边形。

点有一个(X,Y)坐标和一个属性记录。

线可以是线段也可以是弧段,包含一条或一组(可连,可不连)的多义线。一条多义线的本质是一组有序结点。线也有一个属性记录。

多边形包括一个或多个边界,一个边界是一个无交叉点的闭合环,一个边界可嵌于一多边形中而形成环形。边界的方向决定它是否代表区域内的面积。多边形也有相应的属性记录。

这里主要介绍针对Shapefile文件的电子地图开发。

3.1.1 图层控制

通常将图层分为点状图层、线状图层和面状图层,分别对应点、线、面3种地理要素。一幅地图通常是由多个图层叠加产生的,由图层加载的时候要使点状图层在最上面,线状图层在中间,面状图层在最下面。如果需要注记,还要在最上面添加装饰图层。注意在MO中,最先加载的在下面,最后加载的在上面。

目前,很多GIS平台厂商都是以控件的形式发布他们的二次开发产品,这使得我们的

编程实现显得非常容易，我们可以像使用其他 Windows 控件一样使用 MO。一种方法是，直接将控件拖放到窗体上；还有一种方法，先导入类型库，然后使用代码动态创建控件。显然，第二种方法更灵活机动，而且可以控制控件的创建成功与否，第一种方法则更直观。

既然 MO 也是一种控件，我们先在 VB .NET 工具箱中添加 Map 控件，方法是：

1）在.NET 的“工具箱”中的空白处，单击鼠标右键选择“添加/移除项”，如图 3-1 所示。

2）单击之后弹出“自定义工具箱”窗口，将 AxMap 前的方框打勾选中，如图 3-2 所示，单击“确定”后，AxMap 控件就可以添加到.NET 的工具箱中。



图 3-1　选择快捷菜单中的“添加/移除项”

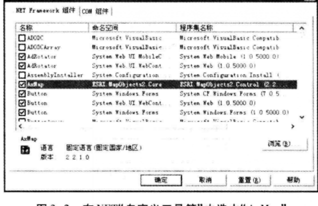


图 3-2　在.NET“自定义工具箱”中选中“AxMap”

3）在工具箱中可以查看 AxMap 控件，如图 3-3 所示。

这样，就可以通过鼠标点选后拖动在一个窗体中添加 AxMap 控件，并且修改属性和布局。和 Windows 中的控件一样，需要为项目中的地图控件定义一个唯一的名称，在属性面板中进行操作即可，本例中的名称为“mapMain”，如图 3-4 所示。

地图控件最基本的功能是显示地图，通常一幅电子地图由若干个图层组成。AxMap 控件支持两种加载显示地图的功能：手工加载图层和程序加载图层。

AxMap 组件提供了程序添加图层的方法，如下面代码所示。该方法的参数，就是需要添加的图层。

```
mapMain. Layers. Add( )
```

手工加载图层是通过设置控件的属性，添加图层文件。

程序加载图层可以利用“dc. Database =‘文件路径’”设置图层文件所在目录，其中，“dc. Database = app. path”为设置与程序存在的文件夹一样的目录。

MO 还为我们提供了更为灵活的图层控制方式，需要借助 legend 控件实现。这样，当在项目中加入或删除图层时，在 legend 中同步更新，由 legend 动态组织和管理图层，控制图层的显示或隐藏。

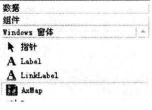


图 3-3　查看“AxMap”控件

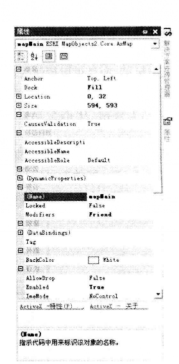


图 3-4　“AxMap”控件的属性

为了使用 legend 控件，首先在项目中添加这个控件，方法是在“工具箱”中单击鼠标右键，选择“添加/移除项”，弹出“自定义工具箱”，在 COM 组件中将 legend 选项前的框里面打勾选中，如图 3-5 所示，然后单击确定。legend 组件就加入到.NET 的工具箱中。

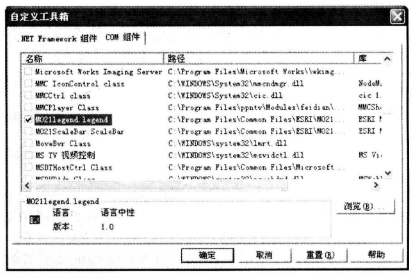


图 3-5　在.NET“自定义工具箱”中选中“legend”

本项目中对图层的控制调用了 MO 中 legend 功能模块，在项目中加入或删除图层时，在 legend 中同步更新。

为项目添加 legend 控件，并命名为“molLayer”，在项目中添加或删除图层的实现代码如下：

```
Me. molLayer. setMapSource( Me. mapMain)   ′* 在 legend 中显示图层信息
Me. molLayer. Loadlegend( True)
Me. molLayer. ShowAlllegend( )
```

在 legend 中可对项目中图层进行调整顺序、显示/隐藏等操作。

3.1.2 地图的放大和缩小

对地图的放大有两种方法：一种是在地图显示区域用鼠标画一个框，然后将框内显示的区域进行放大；另外一种是点“放大”按钮，对地图显示区域当前显示的内容进行放大。

第一种放大功能的实现代码如下：

```
Dim rect As MapObjects2. Rectangle
rect = mapMain. TrackRectangle( )
mapMain. Extent = rect
```

第二种放大功能的实现代码如下。其中，参数 0.6667 表示放大后显示的区域是原来地图显示区域的 2/3。

```
Dim r As MapObjects2. Rectangle = mapMain. Extent
r. ScaleRectangle(0. 6667)
r. Offset( -( r. Center. X-pt. X) , -( r. Center. Y-pt. Y) )
mapMain. Extent = r
```

将地图显示区域进行缩小，实现代码如下。其中，参数表示缩小后项目中显示的地图区域是之前的 1.5 倍。

```
Dim r As MapObjects2. Rectangle
r = mapMain. Extent
r. ScaleRectangle(1. 5)
mapMain. Extent = r
```

3.1.3 全图显示和漫游

如果对地图执行放大或缩小操作之后，又想显示完整的地图区域，实现代码如下。

```
mapMain. Extent = mapMain. FullExtent
```

地图漫游是指，对地图执行放大操作之后，切换地图显示区域的部分。操作方法是将鼠标放置在地图显示区域，鼠标图形会变成小手样，可以拖动鼠标，控制地图显示区域。该功能的实现代码如下。

```
mapMain. Pan( )
```

3.1.4 鹰眼功能

地图显示的鹰眼功能是指，对地图执行放大操作之后，当前地图显示区域出现的部分在整幅地图中的位置可以用矩形框示意。鹰眼功能的效果如图 3-6 所示，可以看出地图显示区域已被放大，当前显示的区域在整幅地图中的位置是右下角鹰眼所示绿色矩形框。

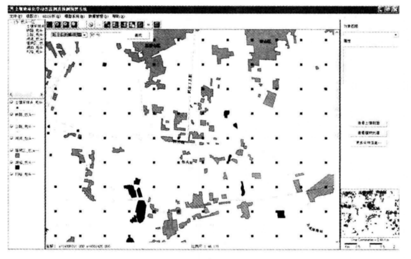


图 3-6 鹰眼功能

鹰眼功能的实现代码如下。

```
mapEye. Layers. Add( Layer)
mapEye. Extent = mapEye. FullExtent
'功能:用新创建的符号在鹰眼窗口中绘制一矩形,矩形大小为地图窗口的显示范围
Private Sub mapEye_AfterTrackingLayerDraw( ByVal sender As Object, ByVal e As ESRI.
MapObjects2. Core. AfterTrackingLayerDrawEventArgs ) Handles mapEye.
AfterTracking LayerDraw
Dim sym As New MapObjects2. Symbol
        sym. OutlineColor = System. Convert. ToUInt32 ( MapObjects2. ColorConstants.
moGreen)
        sym. Outline =True
        sym. SymbolType = MapObjects2. SymbolTypeConstants. moFillSymbol
        sym. Style = MapObjects2. FillStyleConstants. moTransparentFill
'用新创建的符号在鹰眼窗口中绘制一矩形,矩形大小为地图窗口的显示范围
        mapEye. DrawShape( mapMain. Extent, sym)
End Sub
```

3.2 选择功能

地图的选择功能相对比较复杂,MO 提供了点选和框选两种操作。这些功能,可以调用 MO 所提供不同的接口去实现。下面以土壤重金属污染评价及安全预警系统项目中地图选择功能的实现为例,具体介绍实现方法。

3.2.1 点选

表示电子地图里某些图层中的地理要素,可以用下列语句定义:

```
Dim recs As ESRI. MapObjects2. Core. Recordset
```

点选是指,当地图显示区域的地理要素较大的时候,可以直接用鼠标单击该地理要素,表示选中该地理要素,可以查看与该地理要素相关的信息。

定义点选功能,可以用下列语句定义:

```
Dim mapPoint As New MapObjects2. Point  '定义点选
```

执行点选,可以得到鼠标所选中的地理要素的属性,本质上要通过判断鼠标所点击位置的坐标来实现,下面的语句可以实现点选功能。

```
If rect Is Nothing Then
recs = aLayer. SearchByDistance( mapPoint, theTol, "" )
Else
recs = aLayer. Searchshape( rect, 9, "" )
End If
```

要返回选中地理要素的名称，可以用下列语句调用：

```
recs. TableDesc. FieldName( j)
```

3.2.2 框选

框选是指，以鼠标画矩形框的形式，选中框中所有的地理要素，这些地理要素通常是来自不同图层的。如果矩形框画得较大，选择的地理要素会很多，这时可以由用户指定显示某一个或几个图层中的地理要素。

定义框选，可以用下列语句定义：

```
Dim mapRect As New MapObjects2. Rectangle  '定义框选
```

实现框选功能的代码如下。

```
Private Sub Identify( ByVal x As Single, ByVal y As Single, ByVal rect As MapObjects2.
Rectangle)
Dim curCount As Long, layerCount As Long, layer_c As Long
Dim mapRect As New MapObjects2. Rectangle
Dim mapPoint As New MapObjects2. Point
Dim theTol As Double
Dim featCount As Long, fCount As Long
Dim aLayer As Object
Dim recs As ESRI. MapObjects2. Core. Recordset
Dim aName As String, theItem As String
Dim aField As Object
Dim xStr As String, yStr As String
'设置参数
        layer_c = mapMain. Layers. Count
ReDim layerName( layer_c)
ReDim Recs2( layer_c)
        cmbSelected. Items. Clear( )
        ListBox1. ClearSelected( )
```

```
        mapMain. MousePointer = ESRI. MapObjects2. Core. Mouse Pointer
Constants. moHourglass
    mapPoint = mapMain. ToMapPoint(x,y)
    featCount = 0
    layerCount =-1
'坐标处理
If mapPoint. X > 1000 Or mapPoint. Y > 1000 Then
        mapPoint. X = Int(mapPoint. X) : mapPoint. Y = Int(mapPoint. Y)
Else
End If
    xStr = Format(mapPoint. X,"0. 000")
    yStr = Format(mapPoint. Y,"0. 000")
    lblSelectedLocation. Text = "x=" + xStr & ",y=" + yStr
    featCount = 0
    layerCount =-1
'设置误差
    theTol = mapMain. ToMapDistance(SEARCHTOLPIXELS)
'选择对象
Dim k As Integer
    k = mapMain. Layers. Count
Dim curlayers(k-1) As Object
Dim i As Integer
For i = 0 To k-1
        curlayers(i) = mapMain. Layers. Item(i)
Next i
For Each aLayer In curlayers
If aLayer. Visible And aLayer. LayerType = LayerTypeConstants. moMapLayer Then
If rect Is Nothing Then
            recs = aLayer. SearchByDistance(mapPoint,theTol,"")
Else
            recs = aLayer. Searchshape(rect,9,"")
End If
          layerCount = layerCount + 1
          layerName(layerCount) = aLayer. Name
          Recs2(layerCount) = recs
          curCount =-1
If recs. Count <> 0 Then
```

```
                        aName = "Featureid"
Dim l As Integer
                        l = recs. Fields. Count
Dim a(l) As Object
Dim j As Integer
For j = 0 To l-3
                             a(j) = recs. TableDesc. FieldName(j)
Next
For Each aField In a
If recs. Fields. Item ( aField ). Type = ESRI. MapObjects2. Core. FieldTypeConstants.
moString Then
                                  aName = aField
Exit For
End If
Next
End If
While Not recs. EOFReDim Preserve layerNum(2,featCount + 1)
                        curCount = curCount + 1
                        layerNum(1,featCount) = layerCount
                        layerNum(2,featCount) = curCount
                        featCount = featCount + 1
                        theItem = recs. Fields. Item(aName). ValueAsString
If theItem = "" Then
                              cmbSelected. Items. Add(recs. Fields. Item("Featureid").
ValueAsString)
Else
                             cmbSelected. Items. Add(theItem)
End If
                        recs. MoveNext()
End While
Else
                     layerCount = layerCount + 1
End If
Next
If featCount = 0 Then
              lblSelectedNumber. Text = "没有找到任何对象"
              lblSelectedLayer. Text = "对象图层"
              cmbSelected. Items. Clear()
              cmbSelected. ResetText()
              ListBox1. Items. Clear()
```

```
            lblSelectedType. ResetText( )
            lblSelectedLocation. ResetText( )
            lblSelectedNumber. ResetText( )
            mapMain. MousePointer = ESRI. MapObjects2. Core. MousePointerConstants.
moIdentify
Else
            lblSelectedNumber. Text = Str( featCount) + "个对象被找到"
End If
If featCount > 0 Then
            cmbSelected. SelectedIndex = 0
            TextBox2. Visible( ) = False
Else
            lblSelectedNumber. Text = "没有找到任何对象"
End If
        mapMain. Cursor = System. Windows. Forms. Cursors. Arrow
End Sub
```

在土壤重金属污染评价及安全预警系统项目中,选择视图菜单中的"框选"子菜单,或者单击地图常用操作中的"框选"工具按钮,可以打开地理要素框选功能。

在地图显示区域中拖动鼠标,画出一个矩形框,选中一片感兴趣的区域,则该区域中所有的地理要素会闪烁变色3次,并在系统主窗体的右侧上半部分显示所有地理要素的详细属性信息。该功能的执行效果如图3-7所示。

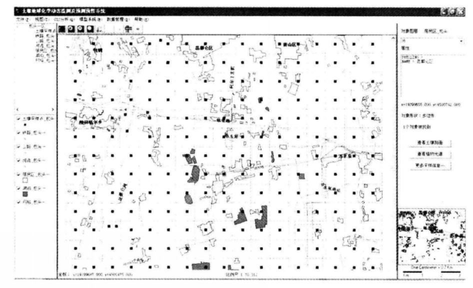


图3-7　框选地理要素

如果找到的地理要素为多个，将逐一列举在地理要素下拉列表框中。用鼠标在列表框中选中一个地理要素，可以在"对象图层"文本框中显示该地理要素所在的图层，在"属性"文本框中显示该地理要素的所有属性。

3.3 查询功能

除了通过地图互操作来实现对地理要素的点选和框选，还可以使用 SQL 语言构建对属性数据的查询。下面以土壤重金属污染评价及安全预警系统项目中的查询功能为例，介绍系统中对不同地理要素属性数据查询的实现。

3.3.1 查询采样点信息

选择"空间分析"子菜单中的"采样点查询"，或者单击地图常用操作中的"查找"工具按钮，可以打开地理要素查找功能，如图 3-8 所示。

在图层下拉列表框中选择要查找的图层，在查找条件文本框中填写查找的条件，单击查找按钮执行查找功能。找到的地理要素会闪烁变色 3 次，并在系统主窗体的右侧上半部分显示该地理要素的详细属性信息。

在 MO 中提供了 SearchExpression 方法，可以按照图层中对象，也就是某一地理要素的名称或者 id 在图层中进行查询。

```
Dim curLayer As MapObjects2. MapLayer
rectSelect = curLayer. SearchExpression( " name = '" + TextBox1. Text + " '" )
```

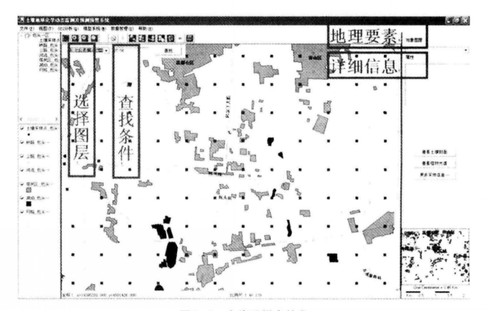


图 3-8 查找采样点信息

例如要查看大气的采样点信息，可以在"选择图层"下拉列表框选择"大气采样点"，整个区域大气采样点的分布情况，如图 3-9 所示。如果需要查询某一采样点具体的信息，可以在"查询条件"文本框中输入采样点的名称即可。还可以在图层控制区域，通过勾选只让地图显示区域显示大气采样点图层，图中每个红色十字形标注代表一个大气采样点，每个小圆圈代表一个土壤采样点。

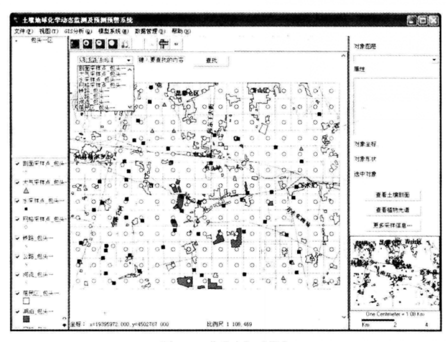


图 3-9　查询大气采样点

3.3.2　查询重金属数据

土壤采样点所采集的土样的重金属有 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等，每种重金属的形态数据包括水溶态、离子交换态、碳酸盐结合态、腐殖质结合态、铁锰氧化物结合态、有机结合态和残渣态等。选择"专题分析"子菜单中的"土壤重金属形态"，可以查询某一采样点土壤中的 8 种重金属含量数据及每种重金属的 7 种形态数据。

3.4　其他功能

地图操作的基本功能还包括距离量算和专题图绘制。

3.4.1　距离量算

距离量算可以帮助我们获得电子地图上任意两个地理要素的距离，方法是定义一个

MO 的“Line”对象,该对象具有“length”方法。

实现距离量算功能的代码如下。

```
Dim line As MapObjects2. Line
line = mapMain. TrackLine( )
Dim length As Double
length = Math. Round( line. Length,3)
MsgBox( length. ToString + " m" , MessageBoxIcon. Information, "距离" )
```

例如,要知道两个采样点之间的距离,可以选择“空间分析”子菜单中的“距离量算”,或者单击地图常用操作中的“量算”工具按钮可以执行地图上任意路线量算的功能。具体方法为:在地图显示区域中单击鼠标,表示量算的起点,拖动鼠标到任一位置再次单击,表示量算的一个折点,双击鼠标表示量算的终点,如图 3-10 所示。

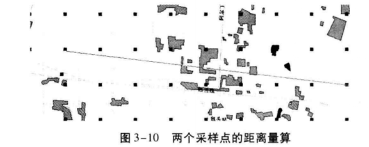


图 3-10　两个采样点的距离量算

弹出的对话框中显示起点到终点所经过的路线的折线距离,如图 3-11 所示。

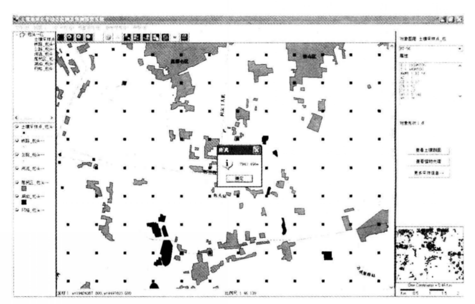


图 3-11　量算结果对话框

3.4.2 坐标与比例尺

MO 提供了坐标显示功能，在程序中定义一个地图中鼠标移动的点，可以将该点的坐标信息取出并显示出来。该功能的代码实现如下。

```
'功能:获取鼠标所指位置地图的坐标
Private Sub mouseLocation(ByVal x,ByVal y)
Dim xStr As String,yStr As String
Dim mapPoint As ESRI. MapObjects2. Core. Point = New ESRI. MapObjects2. Core.
PointClass
        mapPoint =Me. mapMain. ToMapPoint(x,y)
xStr = Format(mapPoint. X,"0.000")
        yStr = Format(mapPoint. Y,"0.000")
        sbpLocation. Text = "坐标: " + "x=" + xStr & ",y=" + yStr
End Sub
'功能:计算地图比例尺
Private Sub mapMain_AfterTrackingLayerDraw(ByVal sender As Object,ByVal e As ESRI.
MapObjects2. Core. AfterTrackingLayerDrawEventArgs) Handles mapMain. AfterTrackingLayer
Draw
        stbLocation. ScaleBarUnits = MO21ScaleBar. sbScaleBarUnits. suKM
        stbLocation. ScreenUnits = MO21ScaleBar. sbScreenUnits. suCentimeters
        stbLocation. PageExtent. MinX = mapMain. Left-21000000/144000
          stbLocation. PageExtent. MaxX = (mapMain. Left + mapMain. Width) -
21000000/144000
          stbLocation. PageExtent. MaxY = (mapMain. Top + mapMain. Height) -
21000000/144000
        stbLocation. PageExtent. MinY = mapMain. Top-21000000/144000
        stbLocation. MapExtent. MaxX = mapMain. Extent. Right
        stbLocation. MapExtent. MinX = mapMain. Extent. Left
        stbLocation. MapExtent. MaxY = mapMain. Extent. Bottom
        stbLocation. MapExtent. MinY = mapMain. Extent. Top
Dim mapScreenWidth As Double
Dim mapExtentWidth As Double
Dim mapScale As Double
'If there are no layers,we cannot calculate scale
If mapMain. Layers. Count = 0 Then
stbLocation. Visible = False
```

```
            sbpScalebar.Text = "比例尺 "
    Exit Sub
    End If
        sbpScalebar.Text = "比例尺 1:" + Format $ (stbLocation.RFScale,"###,###,
###")
        stbLocation.CtlRefresh()
End Sub
```

鼠标移动位置的坐标，以及当前显示地图区域的比例尺，会显示在状态栏里，如图3-12所示。

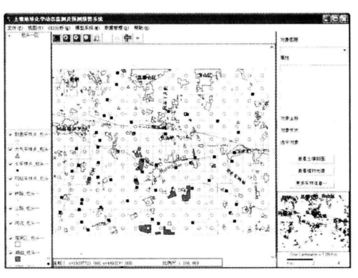


图3-12 坐标与比例尺

3.4.3 专题图

专题地图(Thematic Map)又称特种地图，是地理学中常见的一种图。它是一种着重表示一种或数种自然要素以及社会经济现象的地图。专题地图的内容由两部分构成：专题内容和地理基础。专题内容是指图上突出表示的自然或社会经济现象及其有关特征。地理基础是指用以标明专题要素空间位置与地理背景的普通地图内容，主要有经纬网、水系、境界、居民地等。专题图还具备一般地图的要素，如图例等。

系统的专题图绘制是调用Surfer实现的。土壤重金属污染评价及安全预警系统的专题图分析的内容是重金属含量数据，也就是对所研究区域所采集土壤样本的As、Cd、Cr、Cu、Hg、Ni、Pb、Zn等8种重金属含量数据绘制专题图。

选择"专题图分析"子菜单，执行土壤重金属含量数据的专题图分析，分析结果如图

3-13 所示。

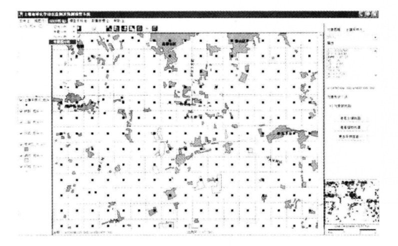


图 3-13　土壤重金属含量数据的专题图

在“重金属种类下拉列表框”中选择一种重金属，则土壤中该重金属含量数据的专题图显示在右侧，如图 3-14 所示。

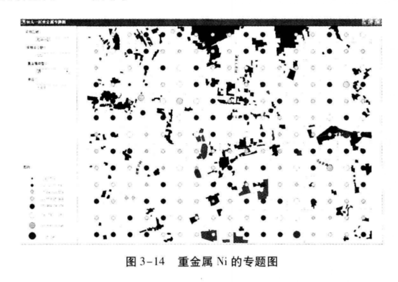


图 3-14　重金属 Ni 的专题图

图中圆点的大小和颜色深浅代表重金属含量的大小，参见左下角专题图的图例即可。同样颜色的圆圈代表重金属含量在一个相同的数值范围，圆圈越大表示含量越高。

本章以土壤重金属污染评价及安全预警系统为例，介绍了 GIS 二次开发的方法，重点是地图操作、地图查询、专题图等功能的实现方法。在 GIS 功能的支持下，有助于对土壤重金属污染数据展开更深入的分析，揭示土壤重金属污染基本情况及空间规律。

第4章 土壤重金属污染数据的统计分析

依据不同的研究视角，土壤重金属污染数据有不同的特性：

从重金属的种类来说，是为了对土壤重金属污染进行客观合理的评价。依据土壤常规检测项目及分析方法的要求，主要考虑砷(As)、镉(Cd)、铬(Cr)、铜(Cu)、汞(Hg)、镍(Ni)、铅(Pb)、锌(Zn)共8种重金属。

从重金属的形态来说，是为研究重金属元素迁移和转化规律。重点应关注的重金属形态包括：水溶态、离子交换态、碳酸盐结合态、腐殖质结合态、铁锰氧化物结合态、有机结合态和残渣态，共7种形态。

从研究土壤采样数据的视角来说，土壤重金属污染数据可以理解为一个高维数据集。有3种度量方法：空间维度上，土壤采样点的地理坐标隐含了土壤重金属污染的空间关系、尺度变异或其他有意义但未知的知识；重金属种类维度上，种类不同对污染程度的贡献不同，且隐含污染源、扩散途径等知识；重金属形态维度上，研究不同重金属形态可以从分子水平、微观角度揭示土壤对重金属的吸附-解析作用及污染的化学机理。

对土壤重金属污染数据的统计分析，就是对区域内的土壤重金属污染数据进行基本的统计分析，如最大值、最小值、平均值、中值、标准差、方差等，并以图表的形式显示。统计图表主要将统计结果以柱状图、饼图、百分比堆积柱状图等方式呈现。

4.1 基本的统计分析方法

4.1.1 基本的统计量

针对土壤重金属污染数据的特性，可采用基本的统计量去分析土样样品中的重金属数据，这些基本的统计量包括最大值、最小值、平均值、中值、方差、标准差等。样本的统计量可以反映样本总体数据的特征。但是实际工作中，由于抽样等原因，样本数据不一定能完全准确地反映总体，可能会与总体存在一定的差异。换言之，不同次的抽样，会得到若干与总体存在差异的统计量。

最大值是样本中所有取值中最大的。最小值是样本中所有取值中最小的。这两个统计量可以表征样本的极大值和极小值，反映样本的取值区间。

平均值反映样本中变量所有取值的集中趋势或平均水平。平均值的计算公式为：

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

其中，n 为有限的样本的个数，x_i 为第 i 个样本。

中值是将样本取值按照顺序排列在一起，这一组数字的中间值。中值往往并不等于平均值。

方差是所有样本值与样本总体平均值的偏差平方的平均值。方差可以表示一组数据分布的离散程度的平均值。方差的平方根可以得到标准差。

方差的计算公式为：

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

标准差的计算公式为：

$$S = \sqrt{S^2}$$

其中，$\bar{x}$ 为样本的平均值，n 为样本的个数，x_i 为第 i 个样本。

4.1.2 调用 Excel 的统计功能

Excel 以单元格为最小的数据存储单位，把数据组织成若干张工作表（Sheet），由若干工作表形成工作簿（Book）。Excel 是常见的重金属微量元素测量仪器形成的检测报告文件形式之一。此外，Excel 具有强大的统计分析功能，提供了“分析工具库”，包括一组数据分析工具，如方差分析、相关分析、协方差、描述统计、指数平滑、F-检验、傅立叶分析、直方图、移动平均、排位与百分比排位、回归分析、抽样分析、t-检验、z-检验等。

每一个分析工具本质上都是由统计函数或工程宏函数来实现的，有些基本的统计函数是内置函数，而其他函数只有在安装了“分析工具库”之后才能使用。

要使用这些工具，可以单击“工具”菜单中的“数据分析”，如图 4-1 所示。如果没有显示“数据分析”命令，则需要加载“分析工具库”加载此项程序。在“数据分析”对话框中，单击要使用的分析工具的名称，再单击“确定”。在已选择的分析工具对话框中，设置所需的分析选项。根据工作簿和用户提供的必要的数据、参数，Excel 的分析工具将调用函数，在输出中显示统计分析结果，其中有些工具在生成输出表格时还能同时生成图表。

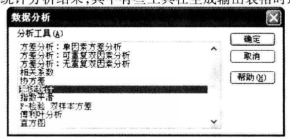


图 4-1 Excel 中的数据分析

在土壤重金属污染评价及安全预警系统中，通过编程调用 Excel 的数据分析功能来实现对土壤重金属数据的统计分析。

Excel 提供了很多统计分析的函数，如下面代码中所示，定义一个 Excel 对象“a”，调用“a”的平均值函数的方法是“a. WorksheetFunction. Average”，方法中的参数“Gycd”即是需要求均值的数组，通常我们需要的数据不需要太多的有效数字，可利用“Math”方法截取有效数字。

要获得具体的一个统计量，就需要分别调用不同的函数。例如，最大值调用 Max 函数、最小值调用 Min 函数、平均值调用 Average 函数、计算中值调用 Median 函数、方差调用 Var 函数、标准差调用 StDev 函数。

相关代码为：

```
Dim a As New Excel. Application
CdA = Math. Round( a. WorksheetFunction. Average( GYcd ) ,3 )
CdMax = Math. Round( a. WorksheetFunction. Max( GYcd ) ,3 )
CdMin = Math. Round( a. WorksheetFunction. Min( GYcd ) ,3 )
CdMedian = Math. Round( a. WorksheetFunction. Median( GYcd ) ,3 )
CdVar = Math. Round( a. WorksheetFunction. Var( GYcd ) ,3 )
CdStDev = Math. Round( a. WorksheetFunction. StDev( GYcd ) ,3 )
```

土壤重金属污染评价及安全预警系统可以分别对数据库中的植物、大气、土壤等样本数据进行基本的描述统计量分析。举例说明，如要执行水样数据的统计分析，可以在打开工作区之后，选择“统计分析”子菜单中的“水样数据”，统计分析的结果如图 4-2 所示。

包头一区水样数据描述统计量

	As	Cd	Cr	Cu	Hg	Ni	Zn	Pb
样本个数：	52	52	52	52	52	52	52	52
均　值：	5.392	0.085	12.356	4.633	5.407	3.737	30.625	0.399
中　值：	2.165	0.03	10.65	2.28	2.245	2.245	6.49	0.16
最大值：	32.139	1.13	41.766	28.678	86.293	15.734	264.882	9.34
最小值：	0.06	0.007	3.64	0.63	0.006	0.009	0.97	0.001
方　差：	41.528	0.028	48.677	37.209	283.466	14.902	3284.98	1.721
标准差：	6.444	0.166	6.977	6.1	16.836	3.86	57.315	1.312

统计结果导出

图 4-2　水样数据的统计分析

要执行土壤数据的统计分析，可以选择“统计分析”子菜单中的“土壤数据”。以包头一地区土壤数据的统计结果为例，首先将包头研究区按照城市功能区划分为工业区、市区、郊区和农村，各区的大致范围如图 4-3 所示。

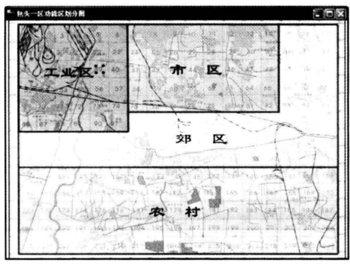
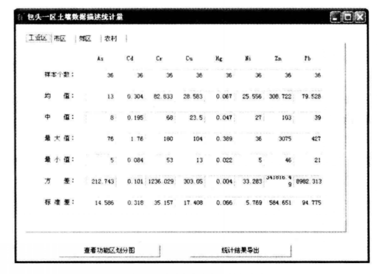


图 4-3　研究区域的城市功能分区

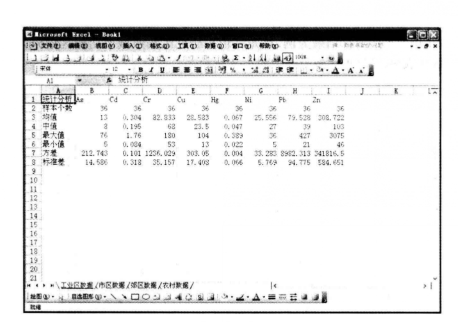
然后分别对各功能区所采集土壤样本数据中 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属的样本进行统计分析，结果如图 4-4 所示。单击工业区、市区、郊区、农村选项卡，可以查看各功能区的统计结果。这些统计结果还可以导出形成 Excel 文件，方法是单击“统计结果导出”，如图 4-5 所示。



	As	Cd	Cr	Cu	Hg	Ni	Zn	Pb
样本个数：	36	36	36	36	36	36	36	36
均　值：	13	0.304	82.833	28.583	0.067	25.556	308.722	79.528
中　值：	8	0.195	68	23.5	0.047	27	103	39
最大值：	76	1.76	180	104	0.389	36	3075	427
最小值：	5	0.084	53	13	0.022	5	46	21
方　差：	212.743	0.101	1236.029	303.05	0.004	33.283	341816.49	8982.313
标准差：	14.586	0.318	35.157	17.408	0.066	5.769	584.651	94.775

图 4-4　土样数据的统计分析

第 4 章　土壤重金属污染数据的统计分析



统计分析	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
样本个数	36	36	36	36	36	36	36	36
均值	13	0.304	82.833	28.583	0.067	25.556	79.528	308.722
中值	8	0.195	68	23.5	0.047	27	39	103
最大值	76	1.76	180	104	0.389	36	427	3075
最小值	5	0.084	53	13	0.022	5	21	46
方差	212.743	0.101	1236.029	303.05	0.004	33.283	8982.313	341816.5
标准差	14.586	0.318	35.157	17.408	0.066	5.769	94.775	584.651

图 4-5　统计结果导出

4.2　常用统计图

对于数据统计分析的结果，除了用基本的统计量描述，还可以用统计图的形式来显示。例如日常生活中常见的人员构成比例图、物品销量变化图、股票交易走势图等，都是以图的形式显示数据。统计图使数据关系体现得更为形象、直观，极大地提高了数据的可视性。

4.2.1　统计图的结构

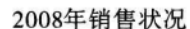
一张统计图可以看成是对数据的图表显示，包括 4 个部分，分别是：统计图标题、图例、绘图区、坐标轴，如图 4-6 所示的某公司两个部门的销售状况统计图。

统计图标题：位于绘图区的上方，说明数据统计图的名称。

绘图区：是图表区中由坐标轴围成的区域。直条、线条、扇面、圆点等都可以作为绘图区的数据标志。同类数据标志的坐标点组成一个数据系列。

图例：为图表中不同的数据系列或分类，指定不同的图案或颜色。图例的位置可以根据图表的情况选择放置在合适的位置。

坐标轴：是度量绘图区图形的参考框架。一般分为垂直坐标轴和水平坐标轴。如果是三维图表，还有第三轴（Z 轴）。坐标轴上可以绘制刻度线，刻度线对应的数字叫做刻度线标签。坐标轴旁边可以加注坐标轴标题。


2008年销售状况

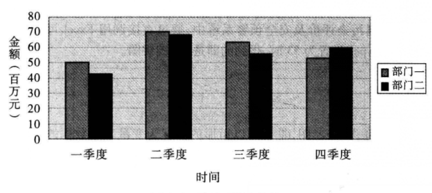


图 4-6　统计图基本结构

4.2.2　Excel 中的统计图

Excel 中提供了强大的统计图绘制功能。统计图比文字、统计量更能直观反映数据的特征，是数据分析不可缺少的部分。在 Excel 中绘制统计图简单、方便，并且有以下优点：

1）功能完善。除了 14 种标准的统计图类型，每种统计图还细分为包含 2 ~ 7 种子图表类型，此外还有 20 种自定义的统计图类型可以套用。各种类型的统计图可以相互组合，这大大扩展了 Excel 绘制统计图的功能。

2）图形美观。用户可以对图表区、绘图区、数据系列、坐标轴、标题、图例、文字颜色、字体、大小、图形背景、图案、颜色等进行修改调整，使得统计图赏心悦目。

3）图与数据关系密切。当统计图和数据放在一张工作表时，改变数据，其统计图将发生相应的变化。将鼠标停留在统计图的某一数据坐标点，在鼠标下方将弹出文本框，给出相应的具体数据。如果用鼠标单击绘图区的“数据系列”标志，则数据系列的单元格会被彩色框线强调显示，便于用户查看图形所对应的数据引用位置。

4）统计图与统计功能联系密切。在“数据系列”单击鼠标右键可以为散点图、线图、条图等添加趋势线，并且可以给出趋势线的方程和系数。

4.3　土壤重金属数据统计图

在土壤重金属污染评价及安全预警系统中，侧重用图的形式来展示重金属数据的特征，主要包含饼图、柱状图、百分比累积图。

4.3.1　饼图

饼图有二维平面式或三维立体式两种，用来显示每一数值相对于总数值的大小。这些数值来自数据表的行或列，行或列中有一定内在关系的数据称之为数据系列。饼图只

有一个数据系列,数据系列中的每个数据项具有唯一的颜色或图案并且在饼图的图例中表示,数据项的大小、各项占总和的百分比例可以标注在饼图的扇面上。此外,对于需要特别强调的部分,可以手动拖出扇面,使之散开以示突出。

在土壤重金属污染评价及安全预警系统中,通过直接调用 Excel 可以绘制饼图,将"activechart. charttype"设置为 53 时,得到的即是 3D 的饼图。

4.3.2 柱状图

柱状图,也叫条图、条状图,是一种以矩形的长度为变量表达数据的统计报告图,由一系列高度不等的纵向或横向条纹表示数据分布的情况。通常使用柱状图来比较两个或以上的数据项。

在土壤重金属污染评价及安全预警系统中,直接调用 Excel 也可以绘制柱状图,将"activechart. charttype"设置为 54 时,得到的即是柱状图,效果如图 4-7 所示。

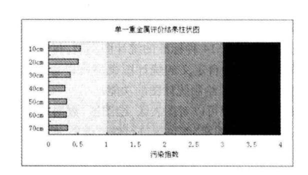




图 4-7 柱状图

4.3.3 百分比累积图

百分比堆积图是一种数据透视方法,用于对大量数据进行分析,其特点是便于查看和比较相交于类别轴上的每一个数值所占总数值的百分比大小。百分比累积图非常适合用于分析重金属形态数据的特性。

要绘制土壤中 8 种重金属(As,Cd,Cr,Cu,Hg,Ni,Pb,Zn)7 种形态(水溶态、离子交换态、碳酸盐结合态、腐殖质结合态、铁锰氧化物结合态、有机结合态、残渣态)的百分比堆积图,所涉及数据量较大,对于每一个土壤采样点,数据的处理方法流程如图 4-8 所示。

在土壤重金属污染评价及安全预警系统中,直接调用 Excel 绘制百分比累积图的功能,将"activechart. charttype"设置为 50 时,可以得到百分比累积图。

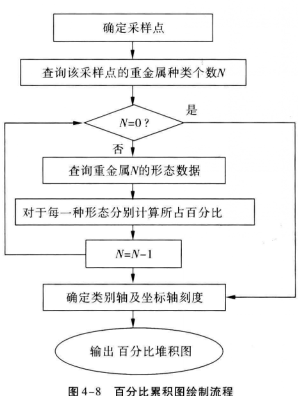


图 4-8　百分比累积图绘制流程

百分比累积图的效果如图 4-9 所示。重金属的种类作为选项卡，每种重金属的 7 种形态分别以饼图显示。“all”选项卡中将所有重金属形态以百分比堆积柱状图输出。

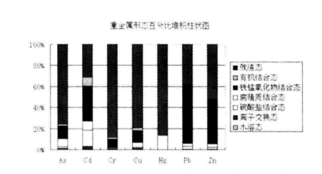


图 4-9　重金属形态统计图

绘制百分比堆积图，对土壤重金属种类、形态数据进行透视分析，实现百分比累积意义上的数据统计分析和挖掘，为重金属微观形态上污染机理研究提供了基础。

第5章 土壤重金属污染数据的空间分析

现实世界中,80%的信息都与地理位置有关,越来越多的地理现象以数字形式表达存储到计算机中,形成地理空间数据库。

空间分析是基于地理对象的位置目标,通过空间计算、空间推理等方法来获取空间信息,求解地理空间问题。常见地理对象的位置目标包括:空间位置、空间形态、空间格局、空间关系等特征信息。

空间分析是基于地理对象空间布局的地理数据分析技术,是地理学和地理信息科学领域的重要研究内容之一,通过空间坐标系统、空间尺度、空间数据结构和空间数据模型等来表达与描述地理现象。对于复杂的地理空间问题可以为其建立空间分析模型,如数字地形模型(Digital Terrain Model,简称DTM)、空间统计分析模型、人工神经网络模型、粗集模型等。目前常用的GIS空间分析方法有缓冲区分析、叠加分析、网络分析、拓扑结构分析、三维分析等。

借助GIS进行土壤重金属污染评价的模型研究和分析,是研究土壤重金属污染系统的重要途径,对于分析重金属污染的空间分布情况、探究污染的本质特征及其动态变化方面具有重要价值。

5.1 等值线和等值面

等值线是将值相等的散点用曲线连接起来,并在每条等值线上标注值。等值面是由等值线形成的图形,利用线的交、并、差运算形成面,并为不同的等值面设置不同的颜色。等值线和等值面的形成离不开数据的插值,通常将离散的数值以内插的方式优化成网格,弥补数据点不足或者分布不均造成的误差。

在土壤重金属污染研究中,土壤样本代表了由点(或小区域)所构成的集合,可以视为空间点分布过程。土壤采样是区域内的离散点,分析土样后得到的重金属污染数据可以作为数据散点,根据这些数据点生成连续表面,通过空间自相关进行空间预测。基于空间点分布形态来研究土壤重金属污染,有助于了解污染的空间分布状况、形成的原因及其可能的扩散趋势。

5.1.1 Kriging 插值

假设最基本的空间点过程为均匀泊松点过程(对丛聚或规则的分布要区别对待),也就是通常所定义的完全空间随机的点分布形态,可以以空间统计理论(也称地统计)来研究区域化变异。该理论是20世纪60年代在法国统计学家G. Matheron的大量理论研究的基础上形成的一门新的统计学分支。

空间统计理论以区域化变量理论为基础,以变差函数为主要工具,研究空间分布上既具有随机性又具有结构性的自然现象的科学。虽然空间数据统计分析是分析空间数据广泛使用的一种方法,但是,必需指出统计分析方法是假设在空间中分布的数据具有统计独立性的,而在现实中,数据由于空间的近邻而具有很大的相关性。

空间自相关统计学的核心和基石是Tobler提出的地理学第一定律:事物彼此关联,但较近的事物比较远的关联性更强。根据空间数据的自相关性,假设研究区中所有的值都是非独立的,相互之间存在相关性。在研究空间分布数据的结构性和随机性、或空间相关性和依赖性、或空间格局与变异、并对这些数据进行最优无偏内插估计,或模拟这些数据的离散性、波动性时,均可应用空间统计学的理论及相应方法。

对土壤重金属污染的空间统计分析包含两个显著的任务:揭示重金属污染的空间相关规律;利用相关规律进行未知点预测。由于插值的直接结果通常不易表示和描述空间规律,一般以等值线和等值面的形式来刻画重金属污染的分布形态,揭示在连续的地理区域中,相邻地理要素及属性的差异。

常见的空间插值方法主要包括以下几种:

(1)距离倒数乘方格网法

距离倒数乘方格网化法是一个加权平均插值法,可以以确切的或者圆滑的方式插值。方次参数控制着权系数如何随着离开一个格网结点距离的增加而下降。对于一个较大的方次,较近的数据点被给定一个较高的权重份额;对于一个较小的方次,权重会比较均匀地分配给各数据点。

(2)最近邻点插值法

最近邻点插值法(Nearest Neighbor)又称泰森多边形方法,是荷兰气象学家A. H. Thiessen提出的一种分析方法。

(3)Kriging法

Kriging法又称空间自协方差最佳插值法,是一种在许多领域都很有用的地质统计格网化方法。Kriging法可以表示隐含在数据中的趋势。Kriging法中包含了几个因子:变化图模型、漂移类型和矿块效应。对土壤重金属污染情况进行空间分析时,多采用Kriging法实现空间插值。

Kriging插值法是空间统计分析方法的重要内容之一,其理论基础是半变异函数理论。半变异函数也称半方差函数,是描述区域化变量随机性和结构性特有的基本手段,其计算方法是:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2$$

式中，$Z(x_i)$和$Z(x_i+h)$表示区域化变量，两者空间位置分别为x_i和x_i+h。$N(h)$是分隔距离为h的样本量。

假设$Z(x)$为区域化变量，并且满足平稳条件和本征假设，可以从数学上证明，半变异函数大时，空间相关性弱。影响半变异函数的主要因素有：样点间的距离的大小、样本数量的多少、异常值、比例及漂移效应等。半变异函数的曲线图如图5-1所示。

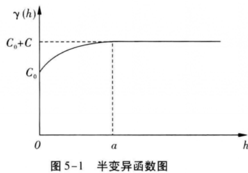


图5-1　半变异函数图

图中，C_0称Kriging方差，表示区域化变量在小于观测尺度时的非连续变异；"C_0+C"为基台值，表示半变异函数随着间距递增到一定程度后出现的平稳值；C为拱高或称结构方差，即基台值与Kriging方差之间的差值；a为变程，是半变异函数达到基台值的间距，通常把变程a视为空间相关的最大间距，也称极限距离。

Kriging法作为一种最优内插法，广泛地应用于地下水模拟、土壤制图等领域，是一种很有用的地质统计格网化方法。它首先考虑的是空间属性在空间位置上的变异分布，确定对一个待插点值有影响的距离范围，然后用此范围内的采样点来估计待插点的属性值。该方法在数学上可对所研究的对象提供一种最佳线性无偏估计（某点处的确定值）的方法。它是考虑了信息样品的形状、大小及与待估计块段相互间的空间位置等几何特征以及品位的空间结构之后，为达到线性、无偏和最小估计方差的估计，而对每一个样品赋予一定的系数，最后进行加权平均来估计块段品位的方法。但它仍是一种光滑的内插方法。在数据点多时，其内插的结果可信度较高。

按照空间场是否存在漂移（drift），可将克里金（Kriging）插值分为普通克里金和泛克里金，其中普通克里金（Ordinary Kriging，简称OK法）常称作局部最优线性无偏估计。所谓线性是指估计值是样本值的线性组合，即加权线性平均；无偏是指理论上估计值的平均值等于实际样本值的平均值，即估计的平均误差为0；最优是指估计的误差方差最小。

Kriging插值法是对有限区域内的区域化变量进行无偏最优估计的一种方法。基于这种方法对数据进行插值时，不仅考虑了未知预测点与邻近样点数据的空间距离关系，还考虑了各参与预测的样点之间的位置关系，充分利用了各散点数据的空间分布结构特征，使其估计结果比传统方法更精确更符合实际，更有效地避免了系统误差的出现。

使用Kriging法插值法处理土壤重金属污染数据的流程如图5-2所示。

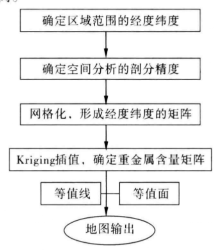


图5-2　土壤重金属污染数据的Kriging插值

5.1.2 调用 Surfer 绘制

Surfer 具有强大的等值线、等值面绘制功能,并且提供了编程接口。在土壤重金属污染评价及安全预警系统中,查询区域的土壤重金属数据,直接调用 Surfer 来绘制等值线、等值面图。

定义一个 surfer 对象,利用 surfer 对象中提供的各种接口来完成等值线的绘制。相关代码为:

```
Dim SurferApp As New Surfer. Application
        SurferApp =CType( CreateObject( "Surfer. Application" ),Surfer. Application)
        SurferApp. Visible =False
         SurferApp. GridData ( DataFile: = InFile, Algorithm: = Surfer. SrfGridAlgorithm.
srfKriging, DupMethod: = Surfer. SrfDupMethod. srfDupNone, ShowReport: = False, OutGrid:
=GridFile)
Dim docu As Object
Dim Plot As Object
        docu = SurferApp. Documents
Plot = SurferApp. Documents. Add( Surfer. SrfDocTypes. srfDocPlot)
Dim shapes As Object
        shapes = Plot. shapes
Dim MapFrame As Object
        MapFrame = shapes. AddContourMap( GridFileName: =GridFile)
Dim ContourMap As Object
        ContourMap = MapFrame. Overlays(1)
ContourMap. ShowColorScale = True
Dim Levels As Object
Levels = ContourMap. Levels
Levels. LoadFile( FileName: ="c:\level. lvl" )
      SurferApp. Documents. SaveAll( False)
        SurferApp. Quit( )
```

上面的代码中,对象“ContourMap. Levels”来定义等值线绘制时的分级情况。

5.1.3 土壤重金属含量等值线

绘制研究区域某种重金属含量的等值线,首先要在数据库中查询区域的重金属污染采样点有哪些,调出每个采样点每种重金属的污染数据,然后进行 Kriging 插值,最后形成

区域重金属含量数值的等值面图。

2D 等值面显示效果如图 5-3 所示，8 种重金属对应着 8 张的等值线，以选项卡的形式查看，等值线间的区域颜色不同，越深代表数值越大。

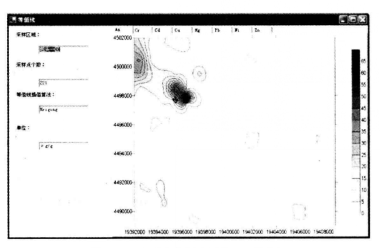


图 5-3　土壤重金属含量的 2D 等值面

5.1.4　土壤重金属含量等值面

为了立体突出显示重金属含量的数据，可以用 3D 立体等值面。同样，对于 8 种重金属以选项卡的形式查看。三维坐标中 Z 值为重金属含量数值，最高的表示最大的，同时颜色较浅，显示效果如图 5-4 所示。

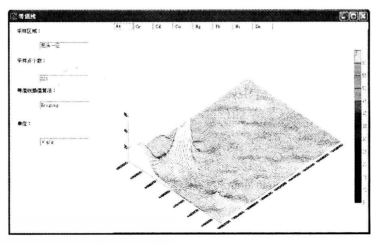


图 5-4　土壤重金属含量的 3D 等值线

在 3D 等值线中，用 45°正视图的方式显示，这样不仅立体感强，便于直观观察重金属

含量极值出现的位置，还可以通过峰谷了解重金属污染的空间变化趋势。

5.2 组件式土壤重金属剖面曲线模拟

为了研究土壤纵深剖面方向重金属污染的详细情况，野外采样时通过钻孔，取得了距离表层一定深度的土壤剖面样品。土壤剖面采样的深度为 2 m，距地表 1 m 的范围内每 20 cm 一个剖面层，距地表 1 ~ 2 m 的范围内每 50 cm 一个剖面层。对每一个剖面层都分析 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属。

为了将土壤剖面各种重金属的含量数据直观显示出来，采用三次样条插值算法来曲线拟合出剖面线，并在 Matlab 中编写成组件供调用。

5.2.1 三次样条曲线拟合

设土壤中某一剖面某种重金属含量满足函数关系 $y=f(x)$ 的列表函数值：

x	x_0	x_1	L	x_n
y	y_0	y_1	L	y_n

可以使用插值方法估计任意的 y 值。多项式函数由于其计算的简单性，在数值近似方面广泛应用。一般来说，分段低次样条插值计算简单、稳定性好、收敛性有保证，且易在电子计算机上实现。

三次样条函数的定义为，设$[a,b]$上有插值节点，$a=x_1<x_2<\cdots<x_n=b$，对应函数值为 $y_1,y_2,\cdots,y_n$。

若函数 $\varphi(x)$ 满足 $\varphi(x_j)=y_j(j=1,2,\cdots,n)$，$\varphi(x)$ 在$[x_j,x_{j+1}](j=1,2,\cdots,n-1)$上都是不高于三次的多项式，当 $\varphi(x)$ 在$[a,b]$具有二阶连续导数。则称 $\varphi(x)$ 为三次样条插值函数。要求 $\varphi(x)$ 只需在每个子区间$[x_j,x_{j+1}]$上确定 1 个三次多项式，设为：

$$\varphi_j(x)=a_jx^3+b_jx^2+c_jx+d_j(j=1,2,\cdots,n-1)$$

其中 a_j,b_j,c_j,d_j 待定，并要使它满足：

$$\varphi(x_j)=y_j\quad(j=1,2,\cdots,n)$$

$$\varphi_j(x_j)=\varphi_{j-1}(x_j)\quad(j=2,\cdots,n-1)$$

$$\varphi'(x_j)=\varphi'_{j-1}(x_j)\quad(j=2,\cdots,n-1)$$

$$\varphi''(x_j)=\varphi''_{j-1}(x_j)\quad(j=2,\cdots,n-1)$$

为了满足三次样条函数，上面给出 $n+3(n-2)=4n-6$ 个条件，需要待定 $4(n-1)$ 个系数，这样看来要唯一确定三次插值函数，还要附加 2 个边界条件。

实际应用中，可以根据具体问题对三次样条插值在端点的状态要求给出附加边界条件。具体就是，通过限定每个三次多项式的一阶和二阶导数，使其在断点处相等，由此可以较好地确定所有内部三次多项式。此外，近似多项式通过这些断点的斜率和曲率是连

续的。然而,第一个和最后一个三次多项式在第一个和最后一个断点以外,没有伴随多项式。因此必须通过其他方法确定其余的约束。

实际上,由于给定了 N 个断点,并添加了 2 个附加条件。这样只需把每个三次多项式列成特殊形式,通过求解 N 个具有 N 个未知系数的方程组,就能确定三次多项式。一般采用稀疏矩阵,方程式能够简明地列出并求解。

与更高次样条相比,三次插值样条只需较少的计算和存储,且较稳定。与低次插值样条相比,三次插值样条在模拟任意形状时显得更灵活。

通过插值算法来弥补采样数量的不足,并根据插值结果拟合生成土壤重金属污染程度的剖面线,为研究剖面中土壤重金属的空间分布特性与迁移规律提供新方法。

5.2.2 组件式混合编程方法

COM 是 Component Object Module 的简称,它是一种通用的对象接口,不论采用什么语言,只要符合这种接口标准,就可以调用。Matlab 组件式编程,就是利用"combuilder"模块,把 Matlab 下的程序做成 COM 组件,供其他语言调用。

组件式混合编程的方法有 3 个步骤。

(1)在 Matlab 下编写 COM 代码

下面以 Matlab R2007a 为例来说明如何在 Matlab 环境下制作 COM 组件。

首先在 Matlab 下新建一个 M 文件,如图 5-5 所示。在其中编写完成功能的 Matlab 代码,在 Matlab 中调试完毕后保存该 M 文件。

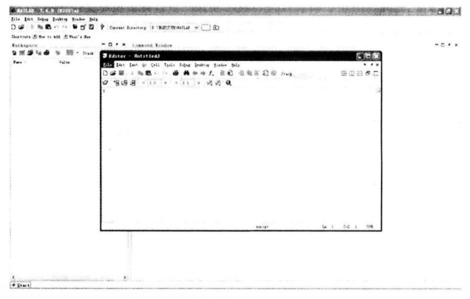


图 5-5 在 Matlab R2007a 中新建 M 文件

在"Command Window"中输入命令"deploytool"进入"Deployment Tool"编辑界面,如图 5-6 所示。

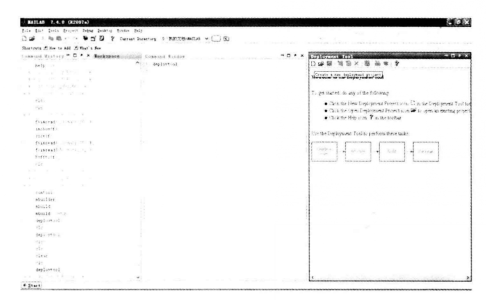


图 5-6 Deployment Tool 编辑窗口

单击新建按钮弹出新建的对话框。在新建对话框中选择“MATLAB builder for.NET”，然后右边出现两个选项，选择“.NET Compnent”，然后命名该项目，并指定存储位置，如图5-7 所示。

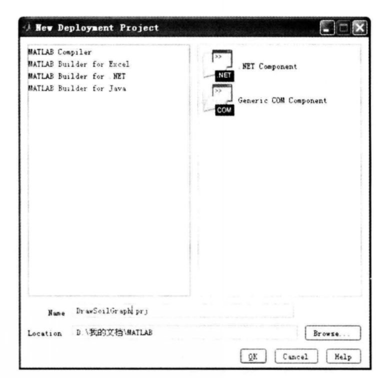


图 5-7 新建一个组件项目工程

点击 OK 按钮，进入组件开发的工作界面，如图 5-8 所示。

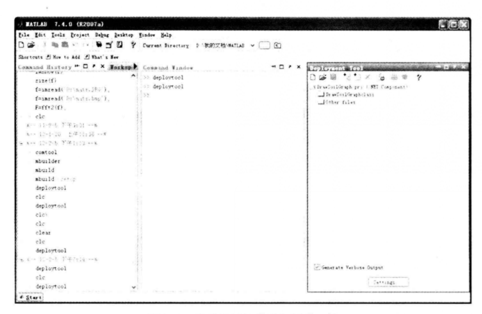


图 5-8　组件项目的编辑和调试环境

在 DrawSoilGraphclass 中添加已经调试好的 M 文件，如图 5-9 所示。在 Matlab 下可以给 COM 组件中的类添加成员、事件、方法，可以参看 Matlab 自带的"combuilder"帮助文件。

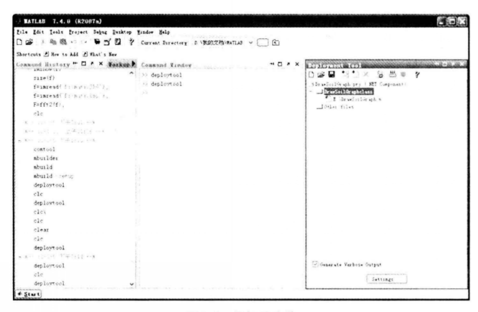


图 5-9　添加 M 文件

(2)编译和打包

然后依次单击按钮和 ，就可以完成 com 组件的制作。

此时到“d:\我的文档\MATLAB\DrawSoilGraph\distrib”文件夹里查看，生成的“DrawSoilGraph_pkg. exe”就是打包后的解压程序，双击它会解压出一些文件，再单击解压出来的“_install. bat”就可以实现安装。

(3)在“VB.NET”中调用组件

要使用制作好的 COM 组件，在“VB.NET”下，为项目添加引用就可以了，如图 5-10 所示。

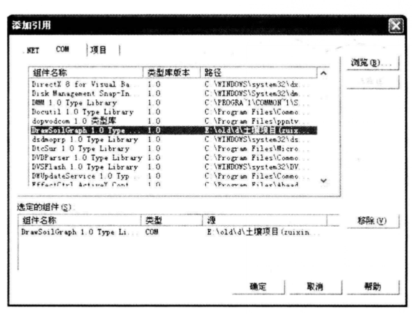


图 5-10　在项目中应用开发的 COM 组件

5.3　土壤剖面线拟合

每个土壤剖面样共有 7 个剖面层，距离地表的深度分布为 20 cm、40 cm、60 cm、80 cm、100 cm、150 cm、200 cm 层，所分析的重金属有 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属，各种重金属分别以曲线拟合出剖面线。土壤剖面曲线拟合的流程图如图 5-11 所示。

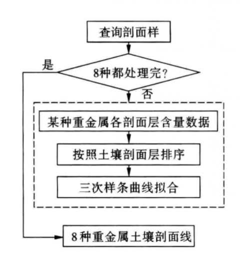


图 5-11　土壤剖面曲线拟合流程

土壤剖面曲线的效果如图 5-12 所示。对于某一个剖面采样点来说,8 种重金属的剖面线分别显示在不同的选项卡中,剖面线上圆圈所示的点为真实的数据。

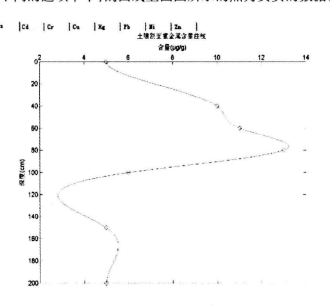


图 5-12　土壤剖面曲线

第6章

基于国标的土壤重金属污染评价

在《中国土壤环境质量标准》(GB 15618—1995)中,有不同种类土壤环境的背景值,当分析重金属污染的程度及空间特性时,可以用来参照。使用这一标准评价土壤环境质量时,可以对单一采样点或某一区域实施评价。对于一个研究区域土壤环境质量的评价,思路是以空间采样为基础,先对多个采样点分别评价,然后根据样本数据进行空间插值,插值处理的本质是加密研究区域的样本数量,从而实现对区域环境质量的整体评价。

本章详细介绍基于GIS的土壤环境质量评价方法,并设计单因子土壤环境质量评价模型、综合因子土壤环境质量评价模型。使用模型不但可以简化评价的计算机实现过程,还可以计算出数值定量评价,也可以以等值线图的形式定性分析区域评价结果。

6.1 重金属污染单因子评价

基于GIS的城市环境土壤质量评价方法将GIS技术引入城市环境质量评价研究之中,达到定性、定量以及定位相结合,从而使评价更加科学化和实用化。

由于地域差异等各种因素的影响,污染指数没有统一的评价标准。国内外学者对土壤环境质量进行评价时,采取的标准也不统一。一般情况下,需要结合区域实际情况来确定研究所采用的评价标准,包括土壤环境质量评价标准或当地的土壤背景值。污染指数法具有一定的客观性和可比性,已在环境质量评价中得到了广泛的应用。

6.1.1 单因子评价流程

单因子土壤环境质量评价是指以某种重金属为主要研究对象,根据区域土壤的背景值,评价该区域内所采土样的污染程度。

单因子评价的流程如图6-1所示,详细步骤为:

1)分析研究区域的特点,确定某种重金属作为重点研究对象,选取该种重金属作为污染因子 F。

2)查询属性数据库,确定研究区域中所有采集土样的个数 N。对于每一个土样分析结果,只找出作为因子的这种重金属的含量。共得到 C_i 个土壤重金属含量数值。

3)对于每一个 C_i,计算单项污染指数 P_i。

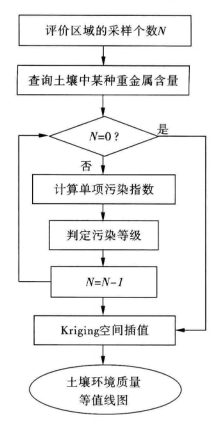


图6-1　单因子评价流程

单项污染指数 P_i 的计算方法为：

若 $C_i \leqslant S_{i污染起始值}$，

则 $P_i = C_i / S_{i污染起始值}$；

若 $S_{i污染起始值} < C_i \leqslant S_{i中度污染起始值}$，

则 $P_i = 1 + (C_i - S_{i污染起始值})/(S_{i中度污染起始值} - S_{i污染起始值})$；

若 $S_{i中度污染起始值} < C_i \leqslant S_{i重度污染起始值}$，

则 $P_i = 2 + (C_i - S_{i中度污染起始值})/(S_{i重度污染起始值} - S_{i中度污染起始值})$；

若 $C_i > S_{i重度污染起始值}$，

则 $P_i = 3 + (C_i - S_{i重度污染起始值})/(S_{i重度污染起始值} - S_{i中度污染起始值})$；

式中 C_i 代表污染物 i 在采样点的实测值，$S_{i\times\times\times}$ 代表污染物 i 不同污染程度的评价标准。

根据《土壤环境质量标准》(GB 15618—1995)中的一级、二级、三级各种重金属参考值，确定不同重金属不同污染程度的起始值。

4)计算得到的 P_i，不仅可以定性污染的程度，还可以根据数量，在每一级的污染程度内细化成不同的污染。

5)根据 N 个 P_i 值，使用 Kriging 空间插值的方法，得到区域的土壤环境质量等值

线图。

需要特别说明的是：

1）从本质上来说，也可以把单因子评价得到的等值线图理解为区域土壤环境某种重金属的污染情况。

2）土壤的背景值可以参照《土壤环境质量标准》（GB 15618—1995），对每个采用评价后再结合 GIS 技术，可以用等值线图的形式给出区域的污染评价结果。

《土壤环境质量标准》是 1995 年制定的，远不能适应当前土壤质量评价的要求。但是，这是国内唯一可以查找到的评价依据，并且该标准给出了土壤中镉、汞、砷、铜、铅、铬、锌、镍等 8 种重金属的自然背景值和污染危害临界值，有重要的参照价值。在实际操作和统一管理方面，依据这一标准中的背景值有重大意义，便于实施，据此可以对全国的土壤做横向的比较研究。

6.1.2 单因子评价模型

通常，对某一研究进行单因子重金属污染评价的时候，采用的方法如前所述。具体操作的时候，可以在 ArcGIS、MapGIS、Surfer 等软件中，根据不同重金属的属性值，在电子地图上叠加新的图层，得到一张等值线图。

但是，如果研究区域的土壤质量比较复杂，或者受哪一种重金属污染最严重不好判断，我们需要一次得到多种重金属的单因子评价结果。事实上，为了对土壤重金属污染进行客观合理的评价，依据土壤常规检测项目及分析方法的要求，《土壤环境质量标准》（GB 15618—1995）中给出了砷（As）、镉（Cd）、铬（Cr）、铜（Cu）、汞（Hg）、镍（Ni）、铅（Pb）、锌（Zn）共 8 种重金属。这 8 种重金属，每个都可以作为一个因子，从土壤受不同重金属污染的角度去分析土壤环境质量。

为了实现"一次得到多种重金属的单因子评价结果"的目标，可以设计区域重金属污染评价的单因子评价模型。将土样中数据各种重金属含量的数据表达成 $m \times n$ 的矩阵，m 代表样本值，n 代表因子。同样，把标准也表达出 $r \times s$ 的矩阵，r 代表土壤环境质量标准（GB 15618—1995）中查到的不同等级对应值，s 代表因子。借助矩阵运算，就可以一次得到多种重金属的单因子评价结果。最后，对于不同的重金属，分别绘制等值线图即可。

土壤重金属污染评价及安全预警系统包含了单因子评价的计算机模型，模型运算后得到一组等值线图，每张等值线图显示在不同的选项卡里，如图 6-2 所示。

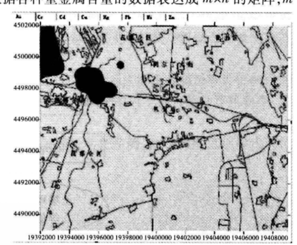


图 6-2　单因子评价模型结果图

从图中可以看到，共有 8 个选项卡，分别是 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn。

6.2 重金属污染内梅罗评价

在实际情况中，常出现多种污染物同时污染某一区域土壤的现象，而单因子污染指数只能反映单个重金属污染物的污染程度，不能全面、综合地反映多种污染物的整体污染水平。因此，设计了国标综合因子模型，即内梅罗评价模型。

6.2.1 内梅罗评价流程

土壤环境质量内梅罗指数评价是指，通过计算综合污染指数来评价土壤环境质量。评价依据是《土壤环境质量标准》(GB 15618—1995)。具体来说，是在单因子评价的基础上，将一个采样的多个重金属单因子评价值合成为一个值。

土壤重金属污染内梅罗指数评价的流程如图 6-3 所示，详细步骤为：

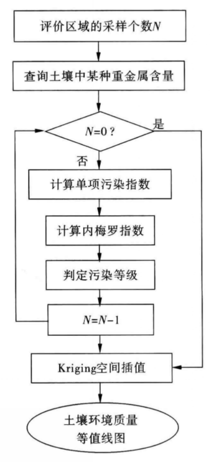


图 6-3 内梅罗评价流程

1）分析研究区域的特点，确定土壤采样的个数 N。

2）查询属性数据库，对于每一个土样分析结果，找到8种重金属的含量。共得到 $8N$ 个土壤重金属含量数值 C_i。

3）根据每一个 C_i，计算单项污染指数 P_i。每个土样有8个单项污染指数，整个研究区域共得到 $8N$ 个单项污染指数。

4）对于每一个土样的8个 P_i，计算内梅罗指数 P_{Nemerow}。每个土样有一个 P_{Nemerow}，研究区域共得到 N 个内梅罗指数。

内梅罗污染指数 P_{Nemerow} 的计算公式为：

$$P_{\text{Nemerow}} = \sqrt{\frac{[\text{average}(P_i)]^2 + [\max(P_i)]^2}{2}}$$

式中 average(P_i) 表示每个土样的8个单项污染指数的平均值；max(P_i) 表示单项污染指数的最大值。

5）根据 P_{Nemerow}，可以确定土样的污染等级。按照《土壤环境质量标准》(GB 15618—1995) 的规定，共分清洁、尚清洁、轻度污染、中度污染、重污染五个等级。

6）根据 N 个土样的 P_{Nemerow} 值，使用 Kriging 空间插值的方法，得到区域的土壤环境质量等值线图。

需要特别说明的是：内梅罗指数抓住土壤中多种污染物的统计特性，即8种重金属污染物单因子评价指数的平均值、最大值，用简单的公式计算合并成为一个值。这种方法，可以反映了8种重金属污染物对土壤的共同作用，并且重点突出了高浓度污染物对土壤环境质量的作用。

6.2.2 内梅罗评价模型

通常，对某一研究进行重金属污染内梅罗综合评价的时候，具体操作可以在 ArcGIS、MapGIS、Surfer 等软件中，根据不同重金属的属性值，先分后合。分是分成多种重金属种类，分别计算单因子污染指数；合是合并单因子污染指数为一。最后，将内梅罗指数作为等值线渲染的 z 值，在电子地图上叠加新的图层，得到一张土壤环境质量综合评价的等值线图。

为了实现"土壤多种重金属污染综合评价"的目标，设计区域重金属污染评价的内梅罗评价模型。土壤重金属污染评价及安全预警系统实现了单因子评价的计算机模型，模型运算后得到一个等值线图，如图6-4所示。

从图中可以看到，研究区域、采样信息、图例都显示在左侧，区域的颜色越深代表污染程度越厉害。

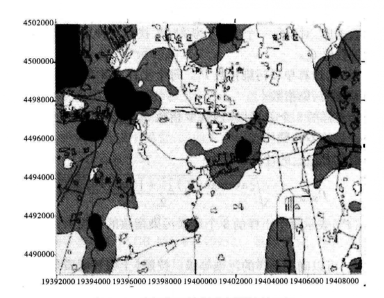


图 6-4　内梅罗污染指数模型评价结果

6.3　评价结果的等值线显示

不管是单因子评价模型，还是内梅罗评价模型，都将模型计算得到的数值利用 Kriging 算法进行空间插值，最后转化成等值线图显示。

这样做的优点是：从地理学的角度对样本数据所反映出来的重金属污染的空间特异性进行分析，利用 GIS 的叠加分析功能，将多种地理要素融合成一张图，为多角度分析污染数据、研究城市重金属污染现状提供参考。对重金属污染的空间分析有助于帮助确定污染源，揭示地理要素、行政区划要素、社会活动要素等同土壤环境污染之间的关系。

第7章 基于神经网络的土壤环境质量评价

影响土壤环境质量的因素很多,基于国标的单因子评价和内梅罗评价用数学解析式来表示重金属含量和污染等级之间的关系,存在一定的不足。从重金属污染过程的角度来说,土壤中重金属种类、形态多样,地域要素差异等多种因素,并且各种因素的重要程度也不尽相同。从重金属评价因子与土壤质量等级间关系的角度来说,两者之间不是简单的函数关系,而是存在复杂的非线性关系,表现出极大的随机性、不确定性和非线性。

人工神经网络简称神经网络,得名于从研究人脑的生理结构出发,模拟人脑处理信息的功能。神经网络是神经科学、数学、统计学、计算机科学、系统工程科学综合的技术,是一种非线性映射方法,构造大量神经元互联的网络,将问题抽象为输入和输出的隐关系,这能够克服人为的主观随意性,获得相对准确有效的运算结果。

基于神经网络的土壤环境质量评价,目标是建立重金属含量和污染等级之间的映射关系,模拟人脑对土壤环境质量评价,并用计算机模型实现。

7.1 神经网络基本原理

神经网络的研究是近年的研究热点,并且在模式识别、时间序列分析、智能控制等领域得到应用。

早在1943年,神经解剖学家和精神病学家 McCulloch 与数学家 Pitts 提出了形式的神经元描述,即 MP 模型,他们所提出神经元遵循"全或无"法则,由神经生理学和数学逻辑学习组成神经网络。MP 模型的提出带动了 NN(Neural Network,神经网络)的研究,同时产生了人工智能学科。1982年,美国生物物理学家 Hopfield 用能量函数构造了具有反馈的神经网络,这是 NN 研究的突破性进展。

神经网络是由大量被称为神经元的处理单元互连而成的,每个神经元是基本的信息处理单元,典型的人工神经元模型如图7-1所示。

图中表示的神经元 i,以 $x_j(j=1,2,\cdots,N)$ 作为输入信号;w_{ij} 作为连接权值;u_i 是输入信号线性组合后的结果,作为神经元的净输入;θ_i 作为神经元的阈值,或称偏差,记为 b_i;$f(\sum_{j=1}^{N} w_{ij}x_j + b_i)$ 作为神经元的激励函数;y_i 作为输出信号。

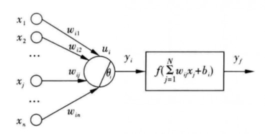


图 7-1 典型的人工神经元模型

加权后的输入可以表示为：

$$u_i = \sum_{j=1}^{N} w_{ij}x_j$$

神经元的输入可以表示为：

$$v_i = u_i + b_i$$

神经元的输出可以表示为：

$$y_i = f(\sum_{j=1}^{N} w_{ij}x_j + b_i)$$

激励函数可以是不同的函数，常见的有 4 种：阈值函数、分段线性函数、Sigmoid 函数和径向函数。

7.1.1 阈值函数

阈值函数的表示形式为：

$$u(t) = \begin{cases} 1 & 若\ t \geqslant 0 \\ 0 & 若\ t < 0 \end{cases}$$

阈值函数也可以看出是阶跃函数，函数图如图 7-2 所示。神经元的输出为 0 或 1，可以理解成是神经元的兴奋或抑制。

7.1.2 分段线性函数

分段线性函数的表示形式为：

$$f(v) = \begin{cases} 1 & 若\ v \geqslant 1 \\ v & 若\ -1 < v < 1 \\ -1 & 若\ v \leqslant -1 \end{cases}$$

分段线性函数的图如图 7-3 所示。特点是在[-1,1]的区间内，对输入的放大是线性一致的，在(-∞,-1)和(1,∞)的区间，对输入的作用可以理解为非线性的。

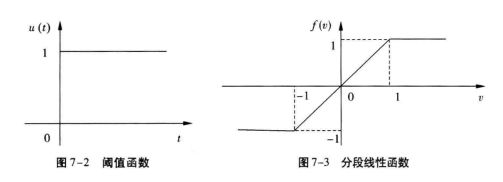


图 7-2　阈值函数　　　图 7-3　分段线性函数

7.1.3　Sigmoid 函数

Sigmoid 函数简称 S 函数。是人工神经网络中最常用的激励函数，它的数学表达形式为：

$$S(t) = \frac{1}{1 + \exp(-at)}$$

式中 exp 表示指数函数，a 称为 S 函数的斜率参数，改变 a 可以获得不同的 S 函数。S 函数的图如图 7-4 所示。当斜率接近无穷时，S 函数转换为简单的阈值函数。

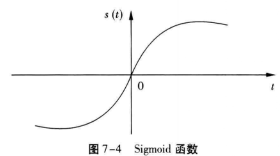


图 7-4　Sigmoid 函数

7.1.4　径向基函数

径向基函数的数学表达形式为：

$$f(t) = \exp\left[\frac{-(t-\mathrm{c})^2}{2\sigma^2}\right]$$

式中 c 是高斯函数的中心，σ 表示方差。径向基函数的图如图 7-5 所示。

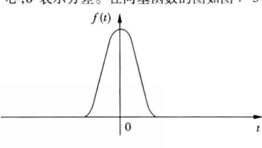


图 7-5　径向基函数

多个神经元可以组成神经网络，神经元之间的连接方式，训练网络的学习方法是神经网络模型的关键。按照不同的角度，可以将神经网络分成多类。例如，按照神经网络结构分为前向网络和后向网络，按照学习方式分为监督学习和非监督学习。不管是那种神经网络，都具有以下三大特性：

（1）非线性

一个神经元可以是线性的，也可以是非线性的。而由多个神经元组成的神经网络具有非线性的特征。通过建立输入和输出之间的非线性映射关系，神经网络能够用来解决许多问题，尤其是为解决输入信号之间可能隐藏非线性物理机制的问题提供了思路。

（2）并行性

神经网络能够解决复杂问题的重要原因之一就是，神经元之间的传递运算是并行的，这就大大提高了运算速度和信息处理能力。

（3）学习性

神经网络具有自学习能力，这种学习可以是监督的，也可以是非监督的。通过对训练样本的学习，神经网络调整神经元的连接权值，达到满足精度要求的目标。由于神经网络的计算过程是一个趋近过程，满足精度的容错是必要的。

典型的神经网络有很多，例如 BP 神经网络，RBF 神经网络，Hopfield 神经网络等。不同的神经网络的网络结构、学习算法各不相同。这里选择 BP 神经网络，构建合适的模型，应用于土壤环境质量评价。

7.2 BP 神经网络模型

BP 神经网络（Back Propagation Netural Network）是一种前馈神经网络，最初的前馈模型是由心理学家 Rosenblatt 于 1958 年提出的。1986 年，Rumehart、Hinton、Willians 提出了一种学习算法，其特性是误差后向传播。目前，对 BP 算法的研究和应用很多，人们通过全面的分析、研究，完善该算法，并在此基础上发展了更多的神经网络模型。

典型的三层 BP 神经网络结构如图 7-6 所示。它由一个输入层、一个输出层和一个或多个隐含层组成。

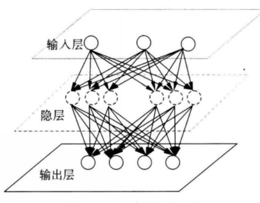


图 7-6　BP 神经网络结构

图中圆形代表神经元，是一个信息处理单元，而且它对信息的处理是相对独立的，若干输入经过一个神经元后得到一个输出。各层次的神经元之间形成全互连连接，各层次内部的神经元之间没有连接，其连接程度用权值表示，并通过学习来调节其值。

BP 神经网络在学习过程中有正向传播和反向传播两部分。正向传播过程为：输入层→隐层→输出层，正向传播是数据由输入层经隐层处理传向输出层；反向传播是误差从输出层向输入层传播并沿途调整各层连接权值和各层神经元的阈值，以使误差不断减小。

BP 算法是基于反向传播算法的网络模型，其本质是建立在梯度下降法基础上的误差修正。BP 算法的基本思想是：

正向传播过程中，每一层的状态只会影响下一层的输入。如果在输出层得不到满足容错要求的输出结果，就转入反向误差传播过程。反向误差传播过程传播的是误差信号，沿原来的连接返回，修改各层的连接权值，使得输出结果的误差满足容错要求。

假设输入层有 M 个输入，输出层有 N 个输出，隐层有 P 个神经元。输入和输出之间的连接是通过神经元拓扑结构完成的。如果用矩阵和向量的形式描述，将输入表示为 $\boldsymbol{X}$，输出表示为 $\boldsymbol{Y}$。

经过输入层神经元，得到：

$$\boldsymbol{X}^{(0)} = f(\boldsymbol{X} - \boldsymbol{\theta}^{(0)})$$

其中，$\boldsymbol{\theta}^{(0)}$ 表示的每个输入层中神经元的阈值所组成的向量。

输入层→隐层过程后，隐层神经元的输入为：

$$\hat{\boldsymbol{X}}^{(1)} = \boldsymbol{W}^{(1)}\boldsymbol{X}^{(0)}$$

其中，$\boldsymbol{W}^{(1)}$ 表示的是输入层→隐层的权值矩阵。

$$\boldsymbol{W}^{(n)} = \begin{bmatrix} w_{11}{}^{(n)} & w_{12}{}^{(n)} & \cdots & w_{1M}{}^{(n)} \\ w_{21}{}^{(n)} & w_{22}{}^{(n)} & \cdots & w_{2M}{}^{(n)} \\ \vdots & \vdots & & \vdots \\ w_{P1}{}^{(n)} & w_{P2}{}^{(n)} & \cdots & w_{PM}{}^{(n)} \end{bmatrix}$$

隐层神经元的输出为：

$$\boldsymbol{X}^{(1)} = f(\hat{\boldsymbol{X}}^{(1)} - \boldsymbol{\theta}^{(1)})$$

其中，$\theta^{(1)}$ 表示每个隐层中神经元的阈值所组成的向量。

输出层神经元的输入为：

$$\hat{\boldsymbol{X}}^{(2)} = \boldsymbol{W}^{(2)}\boldsymbol{X}^{(1)}$$

其中，$\boldsymbol{W}^{(2)}$ 表示的是隐层→输出层的权值矩阵。

$$\boldsymbol{W}_{M1}{}^{(n)} = \begin{bmatrix} w_{11}{}^{(n)} & w_{12}{}^{(n)} & \cdots & w_{1I}{}^{(n)} \\ w_{21}{}^{(n)} & w_{22}{}^{(n)} & \cdots & w_{2I}{}^{(n)} \\ \vdots & \vdots & & \vdots \\ w_{M1}{}^{(n)} & w_{M2}{}^{(n)} & \cdots & w_{MI}{}^{(n)} \end{bmatrix}$$

隐层神经元的输出为：

$$\boldsymbol{Y} = f(\hat{\boldsymbol{X}}^{(2)} - \boldsymbol{\theta}^{(2)})$$

其中,$\theta^{(2)}$ 表示每个输出层中神经元的阈值所组成的向量。

得到的 $\boldsymbol{Y}$ 和期望输出 $\boldsymbol{Z}$ 之间的误差向量可以定义为:

$$\delta = \boldsymbol{Z} - \boldsymbol{Y}$$

BP 算法使用广义误差法则的误差函数,可以定义为:

$$E = \frac{1}{2}\sum_{i=1}^{N}(z_i - y_i)^2$$

修正误差时,注意到权值的改变和多个权值的负微商成正比,对误差函数 E 求 W 和 θ 的导,建立误差梯度矩阵,求导过程不再赘述,最后建立权值的修正计算公式为:

$$\begin{cases}\Delta \boldsymbol{W}^{i}(t+1) = \alpha \boldsymbol{W}^{(i)}(\mathrm{t}) + \Delta \boldsymbol{W}^{(i)}(t) \\ \Delta \boldsymbol{\theta}^{i}(t+1) = \beta \boldsymbol{\theta}^{(i)}(t) + \Delta \boldsymbol{\theta}^{(i)}(t)\end{cases} \quad (i = 1,2)$$

BP 算法的优点是,理论依据鲜明,物理概念清晰,推理过程严谨,算法的实现过程简单可靠。BP 算法是目前比较流行的神经网络学习算法,是能实现映射变换的前馈型网络中最常用的一类网络,是一种典型的误差修正方法。在实际应用过程中,发现 BP 神经网络存在一些缺陷,表现在迭代次数多,收敛速度慢,学习过程长,不一定能保证全局收敛,可能陷入局部最小而得不到整体最优,导致学习过程失败。为了解决应用中存在的问题,可以从以下方面进行改进:改进网络结构,提高收敛速度,全面优化设计,改进传递函数,自适应调整等。

BP 神经网络神经元的传递函数可以采用任意非线性可微非递减函数,在 Matlab 中,提供了多种传递函数,主要有 S 型对数函数(logsig)、纯线性函数(purelin)、S 型的正切函数(tansig)以及它们的导函数。Matlab 的神经网络工具箱还提供对基本的 BP 神经网络的改进,有多种训练函数,如普通的梯度下降训练函数(traingd)、带动量的批处理梯度下降训练函数(traingdm)、利用 Levenberg-Marquardt 算法的训练函数(trainlm)、采用梯度下降动量法的训练函数(traingdx)等。

7.3 BP 神经网络评价土壤环境质量

BP 神经网络是一种非常规非线性的建模方法,由于其强大的非线性映射能力,还具有自组织性、自学习、自适应的特点,可以应用于环境质量评价工作。根据实际应用的需要,设计基于特征模式的 BP 神经网络模型。目标是,建立土壤样本中 8 种重金属含量与土壤环境质量 4 个等级间的非线性关系。

7.3.1 样本选取

由于神经网络特性之一就是学习性,学习过程中训练样本非常重要,BP 神经网络的训练样本的选取对于模型的优劣具有重要的作用,仅通过样本数量的累加和网络的反复训练,不仅耗费大量的时间,而且网络的预测精度得不到实质性的提高。

训练样本的选定,必需考虑整体样本的内在特征和规律。具体到土壤环境质量评价中来说,每个土壤样本分析结果中有 8 种重金属含量,每种重金属含量的数量级不一,对

环境质量的重要性和作用不同。对于一个确定的研究区域，则可以预先根据统计量特征，确定重金属的重要性，据此合理设置训练样本，一般情况下，构造样本的方法可以采用等差序列法建立一批训练样本，但是此种方法不尽合理。如果能更多地考虑样本的统计特性，可以有效地改进训练样本的设置。下面，以包头土壤采样数据的特征及土壤环境质量评价为例，来详细说明训练样本的设置方法。

根据包头的实际情况，222 个土壤采样中，分别对各功能区所采集土壤样本数据中As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属的样本进行统计分析，基本的统计量见表 7-1 所示。

表 7-1　土壤重金属国标背景值与研究区域的统计值　（单位：μg/g）

	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
背景值	15	0.2	90	35	0.15	40	35	100
采样最小值	5	0.051	41	11	0.011	5	16	35
采样最大值	76	1.76	180	104	1.114	41.00	654	3075
采样平均值	7.34	0.19	63.68	20.63	0.08	25.46	38.24	112.89
采样中值	6.00	0.14	60.00	19.00	0.05	25.00	22.00	56.50

神经网络训练样本的输入与土壤中各种重金属的含量相对应，但是从统计结果可以看出，通常 Cd 和 Hg 是小数级，而 Zn 和 Cr 都是十位数级。为了消除各种重金属监测真实数值差异过大对评价造成影响，首先需要对样本的原始数据做规范化处理后再输入神经网络，一种方法是选用单因子评价指数作为规范化处理，单因子评价的计算方法见第五章。通过单因子评价，得到每种重金属的归一化指数，作为神经网络的输入。

神经网络的输出对应土壤质量的评价等级，本研究建立污染特征模式，也就是清洁、轻度、中度、重度污染的特征模式，分别对应向量（1 0 0 0）、（0 1 0 0）、（0 0 1 0）和（0 0 0 1）样本。

从统计结果还发现，包头土壤重金属污染主要有两个特点：一是主要污染物是铅和锌；二是铅和锌伴随出现。单一的设置等差数列样本还不能够准确地评价污染情况，对于一些土壤样本不能正确评价，例如，表 7-2 所示的数据。

表 7-2　土壤重金属污染评价有问题的示例数据

	样本	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn	输出结果			
BT-35	含量/（μg/g）	30.0	0.53	121	26.0	0.18	18.0	182	1457	0.017	0.113	0.674	0.385
	单因子评价	2.33	1.83	1.19	0.71	1.03	0.45	1.47	7.86	中度污染			
BT-113	含量/（μg/g）	5	0.37	74	25	0.19	26	87	596	0.319	0.755	0.194	0.024
	单因子评价	0.33	1.43	0.82	0.71	1.04	0.65	1.16	3.48	轻度污染			

表中数据说明，有些采样点 8 种重金属中有 7 种是清洁的，而 1 种是严重污染，超标

率达百倍以上。表中加粗的评价值表示单因子污染。BT-35、BT-113 中 Zn 的含量高出国标重度污染值上限逾 3 倍,应评为重度污染,但是等差序列样本训练确定的神经网络的评价结果为中度、轻度污染。

为了克服这种问题,需要对神经网络模型进行合理的设置。方法的核心就是,根据统计特性,重点加密某个区间的训练样本。融入统计特性的训练样本的设置采样等差序列法构造,得到的训练样本见表 7-3 和表 7-4。

表 7-3　基于等差序列数值构建的评价训练样本

规范化后的输入样本								输出样本				
As	Cd	Cr	Cu	Hg	Ni	Pb	Zn					
0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35					
0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	1	0	0	0	清洁
0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95					
1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05					
1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	0	1	0	0	轻度
1.65	1.65	1.65	1.65	1.65	1.65	1.65	1.65					
1.95	1.95	1.95	1.5	1.95	1.95	1.95	1.95					
2.05	2.05	2.05	2.05	2.05	2.05	2.05	2.05					
2.35	2.35	2.35	2.35	2.35	2.35	2.35	2.35	0	0	1	0	中度
2.65	2.65	2.65	2.65	2.65	2.65	2.65	2.65					
2.95	2.95	2.95	2.95	2.95	2.95	2.95	2.95					
3.05	3.05	3.05	3.05	3.05	3.05	3.05	3.05					
3.55	3.55	3.55	3.55	3.55	3.55	3.55	3.55					
4.05	4.05	4.05	4.05	4.05	4.05	4.05	4.05	0	0	0	1	重度
4.55	4.55	4.55	4.55	4.55	4.55	4.55	4.55					
5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05					
5.55	5.55	5.55	5.55	5.55	5.55	5.55	5.55					

表 7-3 是设置等差序列构建的训练样本。以 As 为例,清洁的单因子污染指数上限为 1,设置公差为 0.3 的序列,得到 0.35、0.65、0.95 共 3 个样本。轻度污染单因子污染指数为区间数(1,2),设置公差为 0.3 的序列,得到 1.05、1.35、1.65、1.95 共 4 个样本。同样的方法共获得 17 个样本。

单一的设置等差数列样本还不能够准确地评价污染情况,对于类似于表 7-2 所示的

样本不能正确评价，因此增加一些特征样本，如表 7-4 所示。这些样本构造的依据是：参照包头采样的统计特征，依据地区差异设置 Pb、Zn 单因子污染指数较高的等差序列的学习样本。

表 7-4　依据地区差异设置价训练样本

规范化后的输入样本								输出样本				
As	Cd	Cr	Cu	Hg	Ni	Pb	Zn					
0.65	0.65	0.65	0.65	0.65	0.65	1.25	1.25	0	1	0	0	轻度
0.65	0.65	0.65	0.65	0.65	0.65	1.75	1.75					
0.65	0.65	0.65	0.65	0.65	0.65	2.25	2.25	0	0	1	0	中度
0.65	0.65	0.65	0.65	0.65	0.65	2.75	2.75					
1.65	1.65	1.65	1.65	1.65	1.65	2.25	2.25					
1.65	1.65	1.65	1.65	1.65	1.65	2.75	2.75					
0.65	0.65	0.65	0.65	0.65	0.65	3.25	3.25	0	0	0	1	重度
0.65	0.65	0.65	0.65	0.65	0.65	3.75	3.75					
0.65	0.65	0.65	0.65	0.65	0.65	4.55	4.55					
1.65	1.65	1.65	1.65	1.65	1.65	3.25	3.25					
1.65	1.65	1.65	1.65	1.65	1.65	3.75	3.75					
1.65	1.65	1.65	1.65	1.65	1.65	4.55	4.55					
2.65	2.65	2.65	2.65	2.65	2.65	3.25	3.25					
2.65	2.65	2.65	2.65	2.65	2.65	3.75	3.75					
2.65	2.65	2.65	2.65	2.65	2.65	4.55	4.55					

这样设置训练样本，既考虑了所有采样的统计特性，又照顾到了地区差异，训练样本含有两部分：基础样本和特征样本，更为科学合理。

7.3.2 确定网络各层的节点个数和神经元的传递函数

根据前面的建模方法，设置神经网络的输入层节点为8个，输出层节点个数为4个。BP神经网络神经元的传递函数可以采用任意非线性可微非递减函数，主要有S型对数函数(logsig)，纯线性函数(purelin)，S型的正切函数(tansig)以及它们的导函数。在本模型中，隐层神经元的传递函数采用S型正切函数tansig，输出层神经元传递函数采用S型对数函数logsig。经过反复实验，隐层节点数设为13时网络训练时间最短，且能达到最佳的评价效果。

7.3.3 确定训练函数

网络的输入层节点个数为8个，输出层节点个数为4个，经过反复实验，隐层节点数设为13时网络训练时间最短并且能达到最佳的评价效果。模型采用的训练函数是利用Levenberg-Marquardt算法的训练函数(trainlm)。

经反复实验，采用的训练函数是trainlm，运算速度快，能够有效地克服网络训练中的局部最小问题。Matlab神经网络工具箱将所有的学习样本，分成训练样本、确认样本和测试样本3个部分，它们分别占样本总数的1/2、1/4、1/4。经过31步训练后，网络的目标误差小于0.005，达到训练要求，神经网络训练结果如图7-7所示。

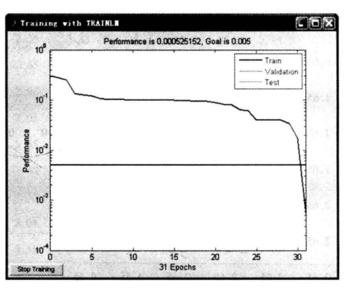


图7-7 神经网络训练结果

7.3.4 模型验证

为了验证模型,将一些通过其它评价方法评价有偏差的例子重新评价,评价结果见表7-5所示。

表7-5　特征值BP神经网络的评价结果

样本		As	Cd	Cr	Cu	Hg	Ni	Pb	Zn	输出结果			
BT-177	含量	6	0.13	54	13	0.02	24	19	38	0.982	0.056	0.013	0.002
	单因子评价	0.40	0.65	0.60	0.87	0.13	0.60	0.86	0.38	清洁			
BT-178	含量	5	0.34	49	16	0.04	19	21	45	0.925	0.437	0.030	0.001
	单因子评价	0.33	1.35	0.54	0.46	0.27	0.60	0.48	0.45	清洁			
BT-179	含量	5	0.18	58	15	0.02	24	17	45	0.937	0.026	0.009	0.001
	单因子评价	0.33	0.90	0.64	0.40	0.13	0.68	0.43	0.45	清洁			
BT-129	含量	5	0.24	51	21	0.11	22	61	310	0.231	0.836	0.327	0.124
	单因子评价	0.33	1.10	0.57	0.60	0.71	0.55	1.08	2.05	轻度污染			
BT-87	含量	5	0.6	57	24	0.06	29	233	515	0.018	0.418	0.761	0.023
	单因子评价	0.33	2	0.63	0.68	0.39	0.73	1.63	3.08	中度污染			
BT-35	含量	30.0	0.53	121	26.0	0.18	18.0	182	1457	0.001	0. 34	0.646	0.928
	单因子评价	2.33	1.83	1.19	0.71	1.03	0.45	1.47	7.86	重度污染			
BT-113	含量	5	0.37	74	25	0.19	26	87	596	0.021	0.205	0.339	0.836
	单因子评价	0.33	1.43	0.82	0.71	1.05	0.65	1.13	3.48	重度污染			

如BT-178点,用内梅罗指数法评价为轻度污染,但是本研究基于特征值的BP神经网络方法评价为清洁的,BT-35和BT-113点,用等差数列为样本的神经网路评价为其他污染等级,这里评价为重度污染。

从模型验证结果来看,基于本研究方法的评价能够正常反映各个采样点的污染情况。样本学习及验证建立好的基于特征值的BP神经网络,可以用来对采样点进行评价,并且能得到较好的评价效果。

第 8 章 基于灰理论的土壤环境质量评价

灰色系统理论,简称灰理论,该理论由邓聚龙教授提出并推广发展。1982 年,邓教授发表论文《灰色控制系统》,标志着灰色系统这一学科诞生。1985 灰色系统研究会成立,灰色系统相关研究发展迅速。1989 年海洋出版社出版了英文版的《灰色系统论文集》,同年,英文版国际刊物《Grey System》杂志正式创刊。

灰色系统理论以灰色关联空间为分析基础,以灰色模型为核心,以灰色过程及其生成空间为研究内容,是一个完善的、系统的方法体系。该理论在很多领域得到了广泛的应用,如工业、农业、社会、经济、能源、地质、石油等众多科学领域,成功地解决了生产、生活和科学研究中的大量实际问题,取得了显著成果。

8.1 灰理论基本原理

灰,是对信息不充分、不完全的一种称谓,与"黑"表示信息缺乏和"白"表示信息完全相对应。灰色系统是指部分信息明确、部分信息不明确的系统,与信息完全明确的系统(白色系统)相对。灰色系统着重外延明确、内涵不明确的对象,模糊数学着重外延不明确、内涵明确的对象。

灰数指只知道大概范围而不知道其确切值的数,灰数是灰色系统理论的基本"单元",最小"细胞"。例如,某人的年龄在 18 岁左右。"18 岁左右"就是一个灰数,这是由于对此人的年龄缺乏充分了解而造成的。如果补充"某人"的出生日期信息,则可以确定其年龄。又如,今天的温度为 15 ~ 20 ℃。"15 ~ 20 ℃"就是一个灰数,这是由于对于温度表述划分不充分造成的。如果补充具体的地点和时间点,那么温度可以确定为一个具体的值。类似地,通过信息补充,可以将灰系统的研究转换为白系统的研究。而模糊数学用模糊集描述不确定的事物,不存在补充信息使命题发生转换。

灰色系统的基本原理有:

公理 1:差异信息原理。"差异"是信息,凡信息必有差异。

公理 2:解的非唯一性原理。信息不完全,不明确的解也是非唯一的。

公理 3:最少信息原理。灰色系统理论的特点是充分开发、利用已有的"最少信息"。

公理 4:认知根据原理。信息是认知的根据。

公理 5:新信息优先原理。新信息对认知的作用大于老信息。

公理6:灰性不灭原理。"信息不完全"是绝对的。

灰色系统理论的主要任务之一是根据社会、经济、生态等系统的行为特征数据,寻找不同系统变量之间或某些系统变量自身的数学关系和变化规律。

灰色系统理论认为,尽管客观系统表象复杂,数据离乱,但它总是有整体功能的,因此必然蕴含某种内在规律。因此,灰色系统理论认为任何随机过程都是在一定幅值范围和一定时区内变化的灰色量,并把随机过程看成灰色过程。

灰色系统理论的本质是通过对原始数据的挖掘、整理来分析数据的变化规律,这是一种就数据寻找数据的现实规律的途径。

灰色系统理论经过数十年的发展,已基本建立起了一门新兴学科的结构体系,其主要内容包括:灰色关联分析、灰色统计、灰色聚类等。该理论为系统分析、评估建模、预测决策、控制优化提供了完整的理论体系和技术实现。

8.2 GRA 模型

灰色关联分析法(GRA)的基本思想是根据序列几何曲线的相似程度来判断其联系是否紧密。曲线越接近,形状越相似,相应序列之间的关联就越大,反之就越小。例如样本数据反映出两个因素变化的态势(方向、大小、速度等)基本一致,则它们之间的关联度较大;反之,关联度较小。与传统的多因素分析方法(相关、回归等)相比,该方法对数据要求较低且计算量小,便于推广应用。

8.2.1 灰色关联因素和关联算子集

灰色关联因素又称灰色关联因子,确定因素集是灰色关联分析的前提,因素状态间的"接近"、"比较"、"辨识"、"评估"是灰色关联分析的基础。将因素作为空间的点,对同一因子在不同指标、不同意义、不同目标、不同方位下的分量记作点的分坐标,并按照距离的定义描述点的接近测度,这样就可以依据距离空间度量灰关联空间。

要对灰色系统进行分析,第一步是选准系统行为特征的映射量,然后进一步明确影响系统行为的有效因素。如要作量化研究分析,则需要对系统行为特征映射量和各有效因素进行处理。使用关联算子集可以完成这个量化过程,通过算子作用,使因素转化为数量级大体相近的无量纲数据,并将负相关因素转化为正相关因素。下面给出一些术语的定义。

(1)初值象

设 $X_i=(x_i(1),x_i(2),\cdots,x_i(n))$ 为因素 X_i 的行为序列,D_1 为序列算子,且 $X_iD_1=(x_i(1)d_1,x_i(2)d_1,\cdots,x_i(n)d_1)$。其中

$$x_i(k)d_1=\frac{x_i(k)}{x_i(1)}\quad x_i((1)\neq 0;k=1,2,\cdots,n)$$

称 D_1 为初值化算子。X_iD_1 为 X_i 在初值化算子 D_1 的象,简称初值象。

(2)均值象

设 $X_i=(x_i(1),x_i(2),\cdots,x_i(n))$ 为因素 X_i 的行为序列，D_2 为序列算子，且 $X_iD_2=(x_i(1)d_2,x_i(2)d_2,\cdots,x_i(n)d_2)$。其中

$$x_i(k)d_2=\frac{x_i(k)}{\frac{1}{n}\sum_{k=1}^{n}x_i(k)}\quad(k=1,2,\cdots,n)$$

称 D_2 为均值化算子。X_iD_2 为 X_i 在均值化算子 D_2 的象，简称均值象。

(3)区间值象

设 $X_i=(x_i(1),x_i(2),\cdots,x_i(n))$ 为因素 X_i 的行为序列，D_3 为序列算子，且 $X_iD_3=(x_i(1)d_3,x_i(2)d_3,\cdots,x_i(n)d_3)$。其中

$$x_i(k)d_3=\frac{x_i(k)-\min_k x_i(k)}{\max_k x_i(k)-\min_k x_i(k)}\quad(k=1,2,\cdots,n)$$

称 D_3 为区间化算子。X_iD_3 为 X_i 在区间化算子 D_3 的象，简称区间值象。

(4)逆化象

设 $X_i=(x_i(1),x_i(2),\cdots,x_i(n))$，$x(k)\in[0,1]$ 为因素 X_i 的行为序列，D_4 为序列算子，且 $X_iD_4=(x_i(1)d_4,x_i(2)d_4,\cdots,x_i(n)d_4)$。其中

$$x_i(k)d_4=1-x_i(k)\quad(k=1,2,\cdots,n)$$

称 D_4 为逆化算子。X_iD_4 为 X_i 在逆化算子 D_4 的象，简称逆化象。

(5)倒数化象

设 $X_i=(x_i(1),x_i(2),\cdots,x_i(n))$，$x(k)\in[0,1]$ 为因素 X_i 的行为序列，D_5 为序列算子，且 $X_iD_5=(x_i(1)d_5,x_i(2)d_5,\cdots,x_i(n)d_5)$。其中

$$x_i(k)d_5=\frac{1}{x_i(k)}\quad(x_i(k)\neq 0;k=1,2,\cdots,n)$$

称 D_5 为逆化算子。X_iD_5 为 X_i 在倒数化算子 D_5 的象，简称倒数化象。

(6)灰色关联算子集

称 $D=\{D_i\mid i=1,2,3,4,5\}$ 为灰色关联算子集。

(7)灰色关联因子空间

设 X 为系统因素集合，D 为灰色关联算子集，称 (X,D) 为灰色关联因子空间。

8.2.2 灰色关联公理与灰色关联度

连续函数可以用欧氏空间的范数作为函数接近的测度，得到的是无限接近的概念。离散函数考虑有限接近，用灰关联度作为一种测度。从本质上来说，灰关联度是分析本征灰系统各种因素间的关联程度。在灰关联空间中，由于这种映射非唯一，只有在确定分辨系数和确定纯量化后才唯一，因此，对于因素间的关联程度，主要用灰关联度的大小顺序描述，而不是用灰关联度的大小描述。

灰色关联度(灰色关联公理)的定义为：

设 $X_0=(x_0(1),x_0(2),\cdots,x_0(n))$ 为系统特征序列,且相关因素序列为:

$$X_1=(x_1(1),x_1(2),\cdots,x_1(n))$$
$$X_i=(x_i(1),x_i(2),\cdots,x_i(n))$$
$$X_m=(x_m(1),x_m(2),\cdots,x_m(n))$$

给定实数 $r(x_0(k),x_i(k))$,若实数 $r(X_0,X_i)=\frac{1}{n}\sum_{k=1}^{n}r(x_0(k),x_i(k))$,满足:

(1)规范性

$0<r(X_0,X_i)\leqslant 1,r(X_0,X_i)=1\Leftarrow X_0=X_i$

(2)整体性

对于 $X_i,X_j\in X=\{X_S\mid s=0,1,\cdots,m;m\geqslant 2\}$,有 $r(X_i,X_j)\neq r(X_j,X_i),i\neq j$。

(3)偶对对称性

对于 $X_i,X_j\in X$,有 $r(X_i,X_j)=r(X_j,X_i)\Leftrightarrow X=\{X_i,X_j\}$ 。

(4)接近性

若 $\mid x_0(k)-x_i(k)\mid$ 越小,则 $r(x_0(k),x_i(k))$ 越大。

满足上述条件的情况,称 $r(X_0,X_i)=\frac{1}{n}\sum_{k=1}^{n}r(x_0(k),x_i(k))$ 为 $X_i,X_j\in X$ 的灰色关联度。其中 $r(x_0(k),x_i(k))$ 为 X_i 和 X_j 在 k 点的关联系数。

上述条件(1)~(4)称为灰色关联四公理。

在灰色关联公理中,规范性表明系统中任何两个行为序列都不可能严格无关联;整体性体现了环境对灰色关联比较的影响,环境不同,灰色关联度也随之变化。因此对称性不一定满足;偶对对称性表明当灰色关联因子集中只有两个序列时,满足对称性;接近性是对关联量化的约束条件。

下面介绍灰色关联分析的步骤。

设有系统行为序列:

$$X_0=(x_0(1),x_0(2),\cdots,x_0(n))$$
$$X_1=(x_1(1),x_1(2),\cdots,x_1(n))$$
$$X_i=(x_i(1),x_i(2),\cdots,x_i(n))$$
$$X_m=(x_m(1),x_m(2),\cdots,x_m(n))$$

对于 $\xi\in(0,1)$,令

$$r(x_0(k),x_i(k))=\frac{\min\limits_i\min\limits_k|x_0(k)-x_i(k)|+\xi\cdot\max\limits_i\max\limits_k|x_0(k)-x_i(k)|}{|x_0(k)-x_i(k)|+\xi\cdot\max\limits_i\max\limits_k|x_0(k)-x_i(k)|}$$

记 $r(x_0(k),x_i(k))$ 为 $r_{0i}(k)$, $r(X_0,X_i)=\frac{1}{n}\sum_{k=1}^{n}r(x_0(k),x_i(k))=\frac{1}{n}\sum_{k=1}^{n}r_{0i}(k)$,则

$r(X_0,X_i)=\frac{1}{n}\sum_{k=1}^{n}r(x_0(k),x_i(k))$ 满足灰色关联公理。

其中 ξ 称为分辨系数。$r(X_0,X_i)$ 称为 X_0、X_i 的灰色关联度,记为 r_{0i} 。

根据关联度的定义和物理意义,可得关联度的计算步骤如下:

步骤1:根据评价目的确定评价指标体系,收集评价数据。

设 m 个数据序列形成如下矩阵:

$$(X_0,X_1\cdots,X_m)=\begin{pmatrix} x_0(1) & x_1(1) & \cdots & x_m(1) \\ x_0(2) & x_1(2) & \cdots & x_m(2) \\ \vdots & \vdots & & \vdots \\ x_0(n) & x_1(n) & \cdots & x_m(n) \end{pmatrix}$$

其中 n 为指标的个数,记 $X_i=(x_i(1),x_i(2),\cdots,x_i(n))^T \quad (i=1,2,\cdots,m)$

步骤2:确定参考数据列 X_0。

参考数据系列应该是一个理想的比较标准,可以以各指标的最优值(或最劣值)构成参考数据列,也可根据评价目的选择其它参照值。

记作 $X_0=(x_0(1),x_0(2),\cdots,x_0(m))$ 。

步骤3:对指标数据序列用关联算子进行无量纲化(也可以不进行无量纲化),无量纲化后的数据序列形成矩阵。

$$\text{记作}\ (X_0',X_1',\cdots,X_m')=\begin{pmatrix} x_0'(1) & x_1'(1) & \cdots & x_m'(1) \\ x_0'(2) & x_1'(2) & \cdots & x_m'(2) \\ \vdots & \vdots & & \vdots \\ x_0'(n) & x_1'(n) & \cdots & x_m'(n) \end{pmatrix}$$

常用的无量纲化方法有均值化象法、初值化象法等。

$$x_i'(k)=\frac{x_i(k)}{\frac{1}{n}\sum_{k=1}^{n}x_i(k)},\quad x_i'(k)=\frac{x_i(k)}{x_i(1)}\quad (i=0,1,\cdots,m;\quad k=1,2,\cdots,n)$$

步骤4:逐个计算每个被评价对象指标序列与参考序列对应元素的绝对差值。

即 $\Delta_i(k)=|x_0'(k)-x_i'(k)| \quad (k=1,\cdots,n;i=1,\cdots,m)$

步骤5:确定极值。

记作 $M=\min\limits_{i=1}^{n}\min\limits_{k=1}^{m}|x_0'(k)-x_i'(k)|$ 与 $m=\max\limits_{i=1}^{n}\max\limits_{k=1}^{m}|x_0'(k)-x_i'(k)|$ 。

步骤6:计算关联系数。

分别计算每个比较序列与参考序列对应元素的关联系数。

记作 $r(x_0'(k),x_i'(k))=\dfrac{m+\xi\cdot M}{\Delta_i(k)+\xi\cdot M} \quad (k=1,\cdots,n)$

式中 ξ 为分辨系数,在(0,1)内取值,ξ 越小,关联系数间的差异越大,区分能力越强,通常 ξ 取0.5。

步骤7:计算关联度。

$$r(X_0,X_i)=\frac{1}{n}\sum_{k=1}^{n}r_{0i}(k)$$

步骤8:依据各观察对象的关联序,得出综合评价结果。

在进行灰色关联分析之前,通常先要对原始数据进行预处理,主要方法如下。

初值化变换：

$$x'_i(k) = x_i(k)/x_i(1)$$

均值化变换：

$$x'_i(k) = x_i(k)/\bar{X}$$

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X(k)$$

百分比变换：

$$x'_i(k) = x_i(k)/\max_k x_i(k)$$

倍数变换：

$$x'_i(k) = x_i(k)/\min_k x_i(k)$$

极差变换：

$$x'_i(k) = \frac{x_i(k) - \min_k x_i(k)}{\max_k x_i(k)}$$

区间值变换：

$$x'_i(k) = \frac{x_i(k) - \min_k x_i(k)}{\max_k x_i(k) - \min_k x_i(k)}$$

8.3 GRA 评价土壤环境质量

灰色关联度评价以"灰色朦胧集"为基础，应用灰色关联分析法将一些边界不清、不易量化的因素定量，是一种综合评价的方法。张先起等将灰色关联分析法用于水质评价中的不确定性研究，同时引入信息熵计算权重。王淑雨等讨论了最大熵模糊综合评价法在污灌区土壤重金属污染等级划分中的应用，该方法计算过程繁琐、计算量大。刘玥等依据国家土壤重金属环境质量分析标准中的 4 种等级，介绍了灰色关联分析法在凤县寺沟铅锌矿区污染评价中的应用，由于只有 5 个监测样本，没有研究细化污染等级的方法。

灰色关联分析法（GRA）的基本思想是根据序列几何曲线的相似程度来判断其联系是否紧密。曲线越接近，形状越相似，相应序列之间的关联就越大，反之就越小。如果样本数据反映出两个因素变化的态势（方向、大小、速度等）基本一致，则它们之间的关联度较大；反之，关联度较小。与传统的多因素分析方法（相关、回归等）相比，该方法对数据要求较低且计算量小，便于广泛应用。因此可以根据影响土壤的因子设置若干评价因子等级标准，然后计算土样和样本的关联度，取关联度最大的等级标准为土样的污染等级。

用灰色关联分析法的土壤重金属污染评价模型进行土壤环境质量评价，具体的流程如图 8-1 所示。

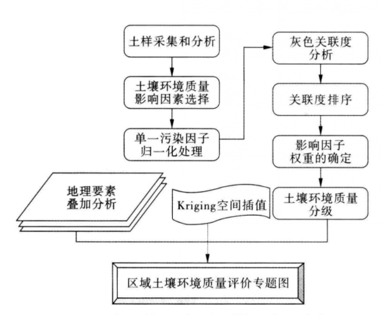


图 8-1 灰色关联分析法的土壤环境质量评价的流程

建立灰色关联分析法的土壤重金属污染评价模型，模型的计算分析步骤如下。

1）原始数据预处理。根据《中国土壤环境质量标准》（GB 15618—1995）中重金属建议范围与数值，计算每种重金属的污染指数。由于评价各指标原始数据量纲不同，数量级差也悬殊，为使各原始数据消除量纲，合并数量级，使其具有可比性，首先对原始数据进行预处理。

设有序列 $X_i=(x_i(1),x_i(2),\cdots,x_i(n))$，经处理之后的序列为 $X'_i=(x'_i(1),x'_i(2),\cdots,x'_i(n))$。在对系统进行分析时，可根据不同的情况选用不同的变换进行原始数据预处理。

2）求差值矩阵。经过预处理的数据序列为：

$$X'_i=(x'_i(1),x'_i(2),\cdots,x'_i(n)) \quad (i=0,1,2,\cdots,m)$$

设评价指标为 X_0，$X_0=(x_0(1),x_0(2),\cdots,x_0(n))$，$\Delta_{0i}(k)=|x_0(k)-x_i(k)|$ （$i=0,1,2,\cdots,m$，$k=0,1,2,\cdots,n$），

可得到差值矩阵，式中 m 表示样本个数，n 表示评价指标个数：

$$\Delta=\begin{pmatrix}\Delta_{01}(1) & \cdots & \Delta_{0n}(1)\\ \vdots & & \vdots\\ \Delta_{01}(m) & \cdots & \Delta_{0n}(m)\end{pmatrix}$$

3）取极大差和极小差。找出差值矩阵中的最大值（极大差）与最小值（极小差），用 Δ_{max} 和 Δ_{min} 表示。

4）求关联系数。

$$\xi_{0i}(k)=\frac{\Delta_{min}+\varepsilon\Delta_{max}}{\Delta_{0i}(k)+\varepsilon\Delta_{max}} \quad (i=0,1,2,\cdots,m;k=0,1,2,\cdots,n)$$

可得到关联系数矩阵：

$$\xi = \begin{pmatrix} \xi_{01}(1) & \cdots & \xi_{0n}(1) \\ \vdots & & \vdots \\ \xi_{01}(m) & \cdots & \xi_{0n}(m) \end{pmatrix}$$

式中 ξ 为分辨系数，$\xi \in [0,1]$。

5）计算关联度。

$$\xi_{0i} = \frac{1}{n}\sum_{k=1}^{n}\xi_{0i}(k) \qquad (i = 0,1,2,\cdots,m)$$

6）排关联序。为准确评价及理顺各子序列对母序列的关联程度，需将关联度依大小顺序排成一列，称为关联序。对于各子序列要比较其对于同一母序列来说孰大孰小，从而可以明确及理顺各子序列对于母序列的"主次"、"优劣"关系。

从统计数据中看出，包头土壤污染中 Pb 和 Zn 的平均值高于背景值，是潜在重要污染因素，而其他 6 种重金属的平均值均小于背景值；另一方面，As、Cd、Pb、Zn 的最大值均超过国标中规定的重度污染的上限，所以这 4 种金属都有可能决定着某个土样污染程度的轻重。

建立土壤重金属污染评价模型时，考虑研究区土壤重金属污染的实际情况，在这里我们选用单因子评价作为规范化处理的方法。参照《土壤环境质量标准》（GB 15618—1995）及主要影响包头地区土壤质量的特征因子，共设置了 8 个特征因子，即：As、Cd、Cr、Cu、Hg、Ni、Pb 和 Zn。

在进行评价之前，先按照单因子评价的结果，对 8 种重金属按照从小到大的顺序进行排序，考虑样本最大值和次大值对污染程度的影响，根据采样的统计特性，将模型的 4 个污染等级细化为 10 类评价指标，样本与哪个指标的关联度大，就将样本评价为相应的污染等级。例如表 8-1 所示的数据，计算关联度之后，与环境质量等级分别对应起来。

表 8-1　GRA 评价土壤环境质量示例

样本编号	关联度										评价结果
	清洁	轻度污染		中度污染			重度污染				
BT-34	0.898	0.693	0.931	0.534	0.693	0.943	0.397	0.484	0.639	0.881	中度污染
BT-87	0.758	0.660	0.773	0.537	0.714	0.815	0.455	0.552	0.718	0.879	重度污染
BT-111	0.949	0.669	0.937	0.467	0.619	0.876	0.358	0.441	0.591	0.842	清洁
BT-177	0.944	0.635	0.900	0.464	0.598	0.856	0.365	0.443	0.574	0.829	清洁
BT-178	0.942	0.640	0.926	0.459	0.592	0.872	0.358	0.434	0.566	0.841	清洁
BT-179	0.940	0.635	0.903	0.461	0.593	0.854	0.362	0.439	0.568	0.825	清洁

第9章

基于数据挖掘的土壤环境质量评价

随着计算机处理数据能力的增强，人们需要从海量数据中挖掘出隐含的规律，以支持信息决策。数据挖掘(Data Mining，又称数据开采、数据采掘)是一种分析、探索数据内在规律的技术，1995年加拿大KDD和DM国际学术会议上由W. J. Frawley和G. Piatetsky等人首次提出。数据挖掘技术植根于计算科学和数学，其成功的关键基石是通过访问正确、完整和集成的数据库，进行深层次的分析。数据挖掘充分利用了机器学习、人工智能、模糊逻辑、分形几何等技术的理论和方法，其研究热点是设计各种高效的算法，以达到从海量数据中发现知识的目的。至今，数据挖掘技术已经在很多领域得到应有，如信息管理、商务分析、图像处理等领域。

地理空间数据挖掘包含旨在从地理空间数据库中发现有用却尚未发现的模式的一系列技术。通过对数据库中的空间数据进行分析与建模，基于空间数据挖掘发现有用的信息，目前已经成为GIS最具生命力的核心功能，这也是GIS区别于其他计算机系统的主要标志之一。由于空间数据库包含了大量的拓扑、距离信息，需要按照复杂的多维空间索引结构组织数据。在访问这些数据时，需要采用空间推理、地理计算和空间知识的表示技术。按照不同的挖掘任务，地理空间数据挖掘可以分为预测模型发现、聚类、关联规则发现、序列模式发现、依赖关系发现、异常值分析和趋势发现等。

空间聚类(Spatial Clustering)是面向空间数据的聚类研究。空间聚类随着空间数据挖掘技术的兴起，已成为地理信息科学和计算机科学共同关注的研究课题之一。目前，地理空间聚类方法主要有四类：分割法、层次法、基于密度的方法及基于网格的方法。聚类和分类重要的不同之处在于：分类需要预先定好类别个数，聚类则是按照设定好的算法自动分类，有没有指定类别个数均可。经典聚类法包括k-means、k-meriod、ISODATA等。近年来，围绕数据挖掘和知识发现领域发展了CLARANS、DBSCAN、Murray等算法。Kohonen自组织特征映射网络、竞争学习网络等自组织神经网络方法，在空间聚类应用中亦有较好的效果。

将数据挖掘技术引入土壤环境质量评价，有以下两个方面的考虑：一是已有的《土壤环境质量标准》(GB 15618—1995)已经不能适应目前环境评价的新形式，背景值的制定对土壤类型、土地利用种类等考虑不够充分；二是需要研究和实现新的高效的计算机评价方法，数据挖掘技术从充分利用已有数据的角度为我们提供了新的思路。

9.1 C 均值聚类算法

聚类分析是机器学习领域的一个重要分支,它参照"物以类聚"的朴素思想,通过研究抽取样本数据的"潜在"结构,将数据对象分组成为多个类或簇,使得同一个簇中的对象之间具有较高的相似度,而不同簇中的对象差别较大。它不需要先验知识和假设,是一种非监督学习方法,广泛应用于数据挖掘和数据分析。

基于目标函数的聚类算法,是把聚类转化成一个带约束的非线性问题,通过优化求解获得对数据的模糊划分和聚类。在这一类方法中,模糊 C 均值聚类算法的理论最为完善,应用最为广泛。

9.1.1 模糊 C 均值聚类算法

模糊逻辑(Fuzzy Logic)理论最早是由 L. A. Zadeh 于 1965 年提出的,是一种描述信息的形式。相对于工程计算、数字推理中精确的测量数字,模糊逻辑理论最大的特点就是用符合人类思维习惯的模糊性的描述去模拟人类的智能活动。例如,高、矮、胖、瘦、年轻、年老等,都是具有模糊逻辑特征的描述。

模糊逻辑理论中,用隶属度函数刻画客观事物由于在形态、类属上的不分明性,描述类似事物间存在的过渡状态,即类别间的不明显的分界线。隶属度函数的值越大,说明隶属程度越高。这反映了人类思维习惯,对于不能明确分解或精确的测量的情况,通常给出模糊性的表述,而不是刻意生硬的划分。

模糊 C 均值聚类算法(Fuzzy C-Means,FCM)将模糊划分的概念引入硬 C 均值算法(Hard C-Means,HCM),其目标函数是由类内平方误差和函数(Within-Groups Sum of Squared Error,WGSS)扩展成的类内加权平均误差和函数。FCM 算法有着深厚的数学基础,建立在正交投影、模糊逻辑和均方逼近的理论基础上,在模糊聚类中有着重要的地位,能够满足数据集量大、实时性要求高的需求。

聚类分析的数学模型可以描述为:对于数据集 $X = \{x_1, x_2, \cdots, x_n\} \subset \mathbf{R}^S$,表示模式空间中 n 个模式的一组有限观测样本集,用 $x_k = (x_{k1}, x_{k2}, \cdots, x_{ks})^{\mathrm{T}} \subset \mathbf{R}^S$ 表示观测样本的特征矢量(或称为模式矢量),x_{kj} 表示特征矢量的第 j 维特征上的取值。

对给定的样本集 X 的 FCM 聚类分析,就是要产生 c 类的划分,满足下面的条件:

$$\begin{cases} X_1 \cup X_2 \cup \cdots \cup X_c = X \\ X_i \cap X_k = \varnothing \ (1 \leqslant i \neq k \leqslant c) \\ X_i \neq \varnothing \text{ 且 } X_i \neq X \ (1 \leqslant i \leqslant c) \end{cases}$$

其中样本 x_k 与子集 $X_i(1 \leqslant i \leqslant c)$ 的隶属关系用模糊集理论的隶属函数 $\mu_{ik} = \mu_{X_i}(x_k)$ 描述。隶属函数的区间为[0,1]。模糊划分得到的是样本隶属于各个类别的不确定性程度,是对类别不确定性的一种度量,可以更客观地反应现实世界。

聚类分析得到的是对样本划分的可能性,为了在众多的分类中寻找一个合理的分类

结果,应该确立合理的聚类准则。最基本的聚类准则是最小平方误差和。

样本 x_k 与第 i 类的聚类原型 p_i 之间的距离度量可以表示为:

$$(d_{ik})^2 = \| x_k - p_i \|_A = (x_k - p_i)^T A (x_k - p_i)$$

其中,$\boldsymbol{A}$ 为是 $s\times s$ 的对称正定矩阵,当 $\boldsymbol{A}$ 取单位矩阵 $\boldsymbol{I}$ 时,与欧氏距离对应。

Bezdek 给出的目标函数可以表示为:

$$\begin{cases} J_m(U,P) = \sum_{k=1}^{n} \sum_{i=1}^{c} (\mu_{ik})^m (d_{ik})^2, m \in [1,\infty) \\ s.t\ U \in M_{fc} \end{cases}$$

其中,$U=[u_{ik}]_{c\times n}$ 表示划分矩阵。$J_m(U,P)$ 表示各类中样本与典型样本的误差平方和。m 表示权重指数,又称平滑参数,其作用可以为隶属度加权。$p_i(i=1,2,\cdots,c)$ 表示第 i 类的典型性代表矢量,或者称为聚类原型矢量,$p_i=(p_{i1},p_{i2},\cdots,p_{is})\in \mathrm{R}^S$。$M_{fc}$ 表示 X 的模糊 c 划分空间,即:

$$M_{fc} = \{ U \in R^{cn} | \mu_{ik} \in [0,1], \forall i,k; \sum_{i=1}^{c} \mu_{ik} = 1, \forall k; 0 < \sum_{k=1}^{n} \mu_{ik} < n, \forall i \}$$

聚类的准则是取 $J_m(U,P)$ 的极小值,使得在满足 $\mu_{ik} \in M_{fc}$ 的约束条件下,聚类结果最优。由于矩阵 U 中各列都是独立的,因此聚类的准则转化为用迭代法求 $J_m(U,P)$ 的极小值。

聚类的准则可以表示为:

$$\min\{J_m(U,P)\} = \min\{ \sum_{k=1}^{n} \sum_{i=1}^{c} (\mu_{ik})^m (d_{ik})^2 \} = \sum_{k=1}^{n} \min\{ \sum_{i=1}^{c} (\mu_{ik})^m (d_{ik})^2 \}$$

极值的约束条件为 $\sum_{i=1}^{c} \mu_{ik} = 1$,使用拉格朗日乘数法求解优化后,得到。

使得 $J_m(U,P)$ 取极小值的 μ_{ik} 的值和此时 p_i 的值分别是:

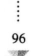
$$\begin{cases} \mu_{ik} = \dfrac{1}{\sum_{j=1}^{c} \left(\dfrac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}}, \text{当 } I_k = \varnothing, \\ \mu_{ik} = 0, \forall i \in \mathrm{I}_{-k} \text{ 且} \sum_{i\in \mathrm{I}_k} \mu_{ik} = 1, \text{当 } I_k \neq \varnothing, \end{cases}$$

$$p_i = \frac{1}{\sum_{k=1}^{n} (\mu_{ik})^m} \sum_{k=1}^{n} (\mu_{ik})^m x_k$$

如果数据集 X、聚类类别数 c、权重值 m 已知,就可以根据上式得到模糊分类矩阵和聚类中心。FCM 算法的步骤为:

1)初始化,给定数据集 X,X 中样本的个数为 n,聚类类别数 c,$2\leq c\leq n$,设置 ε 为迭代停止的阈值。聚类原型的模式 $P^{(0)}$,迭代次数为 b,初始值为 0。

2)根据下式更新矩阵 $\boldsymbol{U}^{(b)}$ 和 $\boldsymbol{P}^{(b+1)}$。



$$
\begin{cases}
\mu_{ik}^{(b)} = \begin{cases} 1, d_{ik}^{(b)} = 0, \\ \dfrac{1}{\sum\limits_{j=1}^{c} \left(\dfrac{d_{ik}^{(b)}}{d_{jk}^{(b)}} \right)^{\frac{2}{m-1}}}, d_{ik}^{(b)} > 0, \end{cases} \\
p_i^{(b+1)} = \dfrac{\sum\limits_{k=1}^{n} \mu_{ik}^{(b+1)} \cdot x_k}{\sum\limits_{k=1}^{n} \mu_{ik}^{(b+1)}}, i = 1,2,\cdots,c
\end{cases}
$$

3)计算 $\| P^{(b)} - P^{(b+1)} \|$,如果小于 ε,则算法停止,否则 $b=b+1$,转向步骤2)。

FCM 算法是目前比较流行的一种模糊聚类算法。在实际应用中存在的问题有:模糊聚类目标函数是非凸的,而 FCM 算法是迭代的,因此很容易陷入局部极值点或者鞍点,得不到最优解或者满意解。此外,FCM 算法是针对特征空间的点集设计的,总是把数据划分到 c 个子类,对于聚类的趋势及有效性分析不充分。

9.1.2 模糊聚类有效性分析

对数据集进行聚类研究,首先应该研究聚类趋势,然后再进行聚类分析,最后要做聚类有效性评估。聚类趋势研究,就是要判断数据集有没有聚类结构。聚类分析研究,就是研究合适的算法,对数据集进行处理,得到聚类结果。聚类有效性分析,就是通过分析聚类结果,判断聚类结果是否合理。

模糊聚类分析的有效性,可以用 3 种方法去判断:基于模糊划分的方法,基于几何结构的方法,基于统计信息的方法。

(1)基于模糊划分的方法

从模糊逻辑理论的角度出发,最后得到的聚类结果应该对数据集有“分明”的划分,或者说隶属度函数的值不能为 0.5,越接近 0 或 1 越好。

为了检验聚类结果的有效性,可以定义分离度、划分熵、基于可能性分布的函数等。这种方法具有简单、运算量小的优点,但是不能定性分析数据集的分类的几何结构特征。

(2)基于几何结构的方法

从数据集划分的几何形态上来说,每个子类的数据集内部数据应当是紧密的,而类别间应当尽可能分离。

为了检验聚类结果的有效性,可以定义划分系数、分离系数、基于图论、Xie-Beni 指标等函数。这种方法运算量大,且表述复杂。

(3)基于统计信息的方法

从分类结果的统计信息出发,对有效的聚类结果应该能得到最佳的统计结果。

为了检验聚类结果的有效性,可以定义熵形式的有效性函数、自举法、PFS 聚类等。

下面简单介绍两种聚类有效性分析函数。

(1)划分系数

对于给定的聚类中心数 c、隶属度矩阵 U,划分系数函数定义为:

$$F(U,c)=\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}{}^{2}$$

其中,n 为待分析的样本数据的个数。

最优的聚类分析结果 c^* 和 U^*,应该满足:

$$F(U^*,c^*)=\max(F(U,c))$$

(2)熵形式

Bezdek 仿照 Shannon 信息熵的形式定义了划分熵的公式:

$$H_m(U,c)=-\frac{1}{n}\sum_{i=1}^{c}\mu_{ik}\cdot\log_a\mu_{ik}$$

熵形式来度量样本间的相近信息,对于给定的 m 值,总希望模糊聚类的划分的熵越小越好。

9.1.3 模糊 C 均值聚类评价土壤环境质量

在众多模糊聚类算法中,模糊 C-均值算法应用最广泛且较成功,它通过优化目标函数得到每个样本点对所有类中心的隶属度,从而决定样本点的类属以达到自动对数据样本进行分类的目的。

土壤样品在经过实验室分析时,获得每个采样点检测土壤中 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属的总量,并且还检测每一种重金属的形态数据,包括水溶态、离子交换态、碳酸盐结合态、腐殖质结合态、铁锰氧化物结合态、有机结合态、残渣态 7 种形态数据。在这里我们使用模糊 C 均值聚类来对单一的重金属的形态数据进行聚类,首先将各种重金属的数据存成 Matlab 可以识别的".dat"格式的文件。然后新建一个 m 文件编写代码。当两次计算目标函数的差的绝对值小于 ε,则算法停止,在本例中进行迭代运算 38 次后满足要求,计算完成。

以 As 为例,将分好的类结合 GIS 技术的专题图形式能更加直观的展示出来,如图 9-1所示,共分 3 类,最大的圆点代表严重污染,次之代表污染,最小的圆点代表清洁。

图 9-1 重金属 As 模糊 C 均值聚类评价结果

模糊集理论及其相关算法在土壤学诸多领域中具有极大的应用潜力，使用模糊 C 均值聚类评价土壤环境质量，是一种新的尝试。但是这种方法在大范围应用推广时，受数据样本空间的限制，有一定的局限性。

9.2 谱聚类算法

传统的聚类方法有 k-means、EM、CURE、DBSCAN 和模糊 C 均值算法等，它们共同的特点是：都建立在基于欧式距离的凸球形的样本空间上，对非凸形的样本空间往往不能达到很好的效果。

谱聚类算法（Spectral Clustering Algorithm）是目前比较流行的一种聚类算法，它能在任意形状样本空间上聚类，且收敛于全局最优解，它的思想来源于谱图划分理论，是由数据点间的相似关系建立矩阵，获取该矩阵的前 n 个特征向量，并且用它们来聚类不同的数据点，该算法对非凸形样本能够取得较好的效果。

9.2.1 谱聚类与空间聚类

谱聚类算法的思想来源于谱图划分理论，如果将数据集看成一个无向完全图 $G=\{V,E\}$，数据点作为图的顶点，将数据点间的相似度量化为定点连接边的权值，则聚类问题就转化为图的划分问题。要使聚类效果达到最优，也就设计一种划分准则，使划分后的子图间的相似度最小，而子图内部的相似度最大。因此划分准则的好坏对聚类效果有直接的影响。目前常用的划分准则主要有规范割准则（Normalized Cut）、比例割准则（Ratio Cut）、平均割准则（Average Cut）、多路规范割准则（Multi-way Normalized Cut）、最大最小割准则（Min-max Cut）以及 Ng，Jordan 等人提出的 Ng-Jordan-Weiss（NJW）等。

NJW 算法的本质是利用相似矩阵的特征向量进行聚类，选取构造矩阵的前 k 个最大特征值对应的特征向量，从而在 k 维空间中构成与原数据一一对应的表述，进而在 k 维空间中利用 k-means 或其他简单算法进行聚类。

NJW 算法的具体步骤如下：

输入：n 个数据点 $S=\{s_1,s_2,\cdots,s_n\}$，聚类数目 K。

输出：数据点集的划分。

1）构造相似矩阵 $A=R^{n\times n}$，其中

$$\begin{cases}A_{ij}=-\exp(-|s_i-s_j|^2/2\sigma^2),i\neq j\\A_{ij}=0,i=j\end{cases}$$

2）构造拉普拉斯矩阵 $\boldsymbol{L}$：

$$\boldsymbol{L}=\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$$

其中 $\boldsymbol{D}$ 是对角矩阵，可以定义为：

$$\boldsymbol{D}=\sum_{j=1}^{n}A_{ij}$$

3）计算矩阵 $\boldsymbol{L}$ 的前 k 个特征值对应的特征向量 $x_1,x_2,\cdots,x_k$，然后构造矩阵$\boldsymbol{X}=[x_1,$

$x_2, \cdots, x_k] \in \mathbf{R}^{n \times k}$,其中特征向量按列存储。

4)对 X 的行向量进行标准化,标准化后的矩阵为 Y:

$$Y_{ij} = \frac{X_{ij}}{(\sum_j X^2{}_{ij})^{1/2}}$$

5)把 $\mathbf{Y}$ 中的每一行看成是 $\mathbf{R}^k$ 空间中样本,然后通过 k-means 算法获得原始数据的最后聚类。

尽管谱聚类和其他聚类方法相比,在解决非凸数据集的聚类问题上具有优势,并在实际应用中取得了很好的效果,但是仍有许多急需研究和解决的问题,主要问题有以下 4 个方面:一是如何构造相似矩阵 W;二是如何选择特征向量;三是如何自动确定聚类数目;四是如何运用到大规模学习问题中。

空间聚类主要用于对海量的空间数据进行深层次分析,揭示出空间数据的分布规律,发现空间离群点,已在地理学、制图学、地质学、遥感学、生物学、经济学等领域得到了广泛的应用。

现有的空间聚类算法大致分为四类:

1)基于划分的聚类算法,具有代表性的算法有 k-means、k-Mediod 等,基于划分的空间聚类算法首先指定空间数据的聚类数目,定义一个目标函数,然后随机选择聚类中心,通过循环迭代,降低目标函数的误差,直到目标函数收敛至一定阈值时,完成聚类。

2)基于层次的聚类算法,如 BIRCH、CURE 等。基于层次的空间聚类算法分为凝聚法和分裂法。凝聚法从每个空间数据实体出发,通过反复聚合,得到不同层次的聚类簇;分裂法对整个空间数据集反复进行分裂,直至所有数据被分裂为单目标的簇,从而得到不同层次的聚类簇。层次聚类算法采用固定的分裂或聚合度量阈值,实质上假设了空间实体分布的均匀性。

3)基于密度和距离的聚类算法,如 DBSCAN、DENCLUE 等。基于密度的聚类方法设定一个阈值,将局部密度大于给定阈值的实体聚为一类;基于距离的聚类算法将实体间空间距离和非空间属性距离均小于一定阈值的实体聚为一类,存在阈值不易确定且很难适应空间分布不均匀情况下聚类的缺陷。

4)基于图论的聚类算法,如 ZEMST 等。基于图论的聚类算法首先在全部数据集内构建一个完全图,每个实体视为图的一个顶点,继而通过打断图的不一致边,形成一系列的子图,每个子图即视为一个簇。然而当空间分布不均匀时,不一致边很难确定。

上述的空间聚类方法都各有利弊。如果能将谱聚类引入到空间聚类中,再对空间对象进行类别归属判定,充分考虑对象与该类别汇总其他对象的空间邻接关系,消除同一类对象出现空间不相邻的现象,进而揭示不同类别对象的空间分布格局。

9.2.2 多维数据聚集趋势检验

对于多维数据而言,特征空间的点与点之间的距离为我们研究数据点的分布特性提供了丰富的信息,但是当数据量较大时,逐对求出点点距离计算量非常大。因此,引入空

间抽样原理,基于抽样的空间统计量结果,对数据集进行聚集趋势检验。

空间统计学的理论基础是空间点过程,指欧几里德空间中的有穷点集合 S 中的点的个数,$N(S)$ 是一个随机变数,是点中有界域 S 内的可数集。最简单、基本的空间点过程 Poisson 过程可以描述为:

$$P\{N(S) = n\} = \frac{e^{-\lambda} \cdot \mu^{n}}{n!}$$

其中,λ 是过程强度,即单位面积或体积(S 在二维空间中具有面积的量纲,在三维及以上空间中具有体积的量纲)内的点的平均数。$\mu = \lambda V(S)$,$V(S)$ 表示有界集合 S 在空间中的体积。如果 $\{S\}$ 是一个空间不相交集合,$N(S)$ 是一个独立随机变量的集合。

模拟 Poisson 过程的方法之一是以 $N(S) = n$ 和上式为条件,在有界域 S 内设置 n 个来自均匀分布的点。这样,对于 S 中的每一个点 x,都有概率密度函数:

$$p(x) = \frac{1}{V(S)}$$

通常,一个理想的聚类应该是点在空间中形成一个椭球体,是一个 Neyman-Scott 聚类过程或 Gauss 分布的混合。典型的 Neyman-Scott 聚类过程包括一个主过程和一个子过程,主过程是一个空间 Poisson 过程,用来产生聚类中心;子过程由两个分布函数组成:离散的 Poisson 分布和径向分布。离散的 Poisson 分布决定产生在单个聚类中的点数,径向分布决定点在空间中相对于聚类中心的距离。

一般 Neyman-Scott 过程的实现,将假定径向分布函数为多元正态分布,满足:

$$f(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \ |C|^{\frac{1}{2}}} \exp[(x - m)^{T} C^{-1} (x - m)]$$

其中,x 为 d 维模式向量,m 为 d 维均值向量,C 为 $d \times d$ 阶的协方差矩阵。

以 m 为中心的多元正态模式群体中空间的椭球体积为:

$$V = \frac{\pi^{\frac{d}{2}}}{\Gamma(0.5d + 1)} \ |C|^{\frac{1}{2}} r^{d}$$

其中,$r = \sqrt{(x - m)^{T} C^{-1} (x - m)}$,$\Gamma(0.5d + 1)$ 是一个伽玛函数。

对基于抽样的空间统计量,进行显著性检验后,得到的检验大小,可以判断数据集是否有聚集趋势。

常用的空间统计量有 4 个,都是基于距离信息的检验统计量:①Hopkins 统计量;②Cox-Lewis 统计量;③T 平方统计量;④Eberhardt 统计量。

上述 4 个统计量的详细描述,可以查数学手册,对于多维数据是否具有聚集趋势有良好的检验性能,但是直接应用存在抽样窗口设置的问题,如何做到随机设置抽样始点和随机选择模式点比较麻烦,如果抽样窗口设置不合理,将直接影响检验统计量。此外,这 4 个统计量适用于均匀分布、规律分布的空间结构,不能用于辨识单个 Gauss 分布的数据集的聚类趋势。

9.2.3 谱聚类评价土壤环境质量

为了表示研究区域内空间对象的邻近关系,我们引入了邻接关系矩阵的概念,表述如下:

设研究区域 S 有 n 个空间对象 $S=\{s_1,s_2,\cdots,s_n\}$，用空间邻接矩阵 W 表达邻接关系，当且仅当空间对象 s_i 和 s_j 具有邻接关系且 $i \neq j$ 时，$W_{ij}=W_{ji}=1$，否则 $W_{ij}=0$。

每个空间对象 s_i 的 m 维属性向量为 $x_i=x(s_i)=[x_{i1},x_{i2},\cdots,x_{im}]$。

考虑空间分布格局的谱聚类算法步骤如下：

1）利用 GIS 软件 ArcGIS 构建空间邻接矩阵 W，并指定谱聚类的聚类数目 K；

2）进行 NJW 算法的前四个步骤；

3）从 Y 中随机挑选 K 个对象，作为初始聚类的各类别中心 $\{z_1(0),z_2(0),\cdots,z_K(0)\}$；

4）将各类别初始化为空，Z_l 初始化为空，执行迭代找出最佳聚类中心，在迭代过程中，如果空间对象 y_j 与其距离最短的集合 Z_l 中的空间对象有邻接关系时，进行 k-means 的下一步；否则下次搜索时不包括该距离，重新进行搜索。

5）若集合 Y 中的所有元素均属于 K 个不同的类别中时，更新个聚类中心值；

6）若所有的聚类中心均保持稳定，即对 $l=1,2,\cdots,K$，有 $z_l(a)=z_l(a+1)$，则 k-means 聚类过程结束，否则重新迭代。

为了验证基于空间聚类的谱聚类算法，以重金属的形态数据为研究对象。以有的研究，很少有直接对各种形态数据和土壤环境质量的关系进行研究，而不可回避的是，重金属的形态决定了其活性及毒性的大小，和污染程度及扩散速度有一定的关系。尝试通过数据挖掘技术，寻找内在规律。

土壤样品经过实验室分析，获得每个采样点检测土壤中 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属的总量，并且还检测每一种重金属的形态数据，包括水溶态、离子交换态、碳酸盐结合态、腐殖质结合态、铁锰氧化物结合态、有机结合态、残渣态 7 种形态数据。表 9-1 为研究区域某一采样点重金属含量数据的示例。

表 9-1　单个采样点重金属含量数据示例　（单位：μg/g）

重金属名称	水溶态	离子交换态	碳酸盐结合态	腐殖质结合态	铁锰氧化物结合态	有机结合态	残渣态	总量
As	0.011	0.001	0.047	2.215	1.401	0.106	6.39	10.17
Cd	0.31	0.3	0.3	140	66.7	3.46	47.1	0.26
Cr	0.054	0.55	0.506	1.687	1.41	4.083	61.35	69.64
Cu	0.07	0.07	0.001	10.74	9.116	0.925	13.49	34.42
Hg	0.64	0.12	0.67	9.83	0.01	0.31	292.6	304.2
Pb	0.014	0.596	1.225	3.259	16.98	1.524	15.8	39.4
Ni	1	1	1	44	7.4	12.9	110	178.3
Zn	0.128	2.642	0.728	6.33	10.26	9.58	112	141.6

由于重金属元素 Pb 的毒性强，对环境影响严重，本文以 Pb 为例，给出基于谱聚类的分析结果。为了使聚类结果更加形象，结合 Kriging 插值和 GIS 可视化分析对结果进行展示，如图 9-2 所示。从图中可以看出，聚类结果为三类。

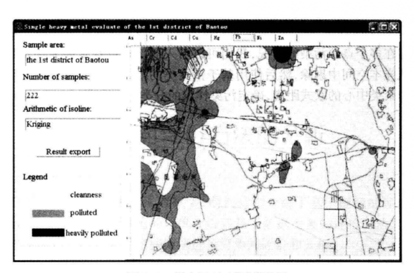


图 9-2　重金属 Pb 谱聚类结果

采用基于谱聚类的空间聚类方法得到的重金属 Pb 的聚类结果，如图 9-3 所示。

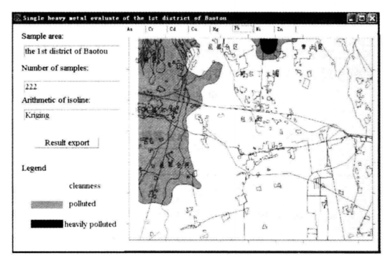


图 9-3　重金属 Pb 考虑空间分布格局的谱聚类结果

从聚类结果图中，可以看出，仅用单一的谱聚类算法，聚类结果存在同一类的对象在空间上处于不相邻的现象，在包头市的郊区和农村存在三个污染点，在工业区包兰线附近出现一个严重污染点，在青山区和昆都仑区之间处在一片严重污染区域。野外实地考察的结果是，包头地区因为包钢的存在，重金属 Pb 的污染基本上是沿昆都伦河呈现条带分布，青山区和昆都仑区交接地带多年前发生过一起加油站含铅汽油泄漏事故，导致土壤中铅的富集，最高值达到 645 mg/g。采样区域内不存在其他异常点源污染区域。考虑空间格局的新的空间聚类算法分析结果，与现状是一致的。

在不存在多个点源污染的情况下，基于谱聚类的空间聚类方法不仅刻画了重金属污

染数据的相似程度，而且还刻画了各类指标的空间分布格局，聚类结果更接近实际的污染情况。

对于存在多个点源污染的情况，我们可以引入局部离群指数的方法，将污染样点作为离群点从样本空间中剔除，然后使用基于谱聚类的空间聚类方法，最后再计算污染样点距离几个聚类中心的欧式距离，确定污染样点的分类。

第 10 章 对评价方法的分析

对于土壤环境质量评价的方法有很多,前面分别介绍了基于国标的评价方法、基于神经网络的评价方法、基于灰理论的评价方法和基于数据挖掘的评价方法。本章对这些评价方法的应用进行介绍,并对这些方法进行简单的评价和分析。

10.1 对内梅罗指数方法的分析

内梅罗指数法兼顾了污染指数的平均值和最高值,在实际应用中,由于此法适应重金属污染因子个数的增减,在加权过程中一定程度上避免了权系数确定过程可能存在的主观因素,是目前应用较多的一种环境质量指数。但是,该方法将复杂的土壤重金属污染评价化为简单的指数形式,过分强调了最高值对环境质量的影响。

例如表 10-1 中出现的情况,3 个采样点的重金属含量相近,但是按照内梅罗法评价结果差异较大。土样 1 被评价为轻度污染,土样 2、3 为清洁。

表 10-1　土壤重金属污染内梅罗评价问题示例

样本		As	Cd	Cr	Cu	Hg	Ni	Pb	Zn	标准差	$P_{Nemerow}$
BT-177	含量/(μg/g)	6	0.13	54	13	0.02	24	19	38	—	0.732
	单因子评价	0.40	0.65	0.60	0.87	0.13	0.60	0.86	0.38	0.251	
BT-178	含量/(μg/g)	5	0.24	49	16	0.04	19	21	45	—	0.863
	单因子评价	0.33	1.10	0.54	0.46	0.27	0.60	0.48	0.45	0.254	
BT-179	含量/(μg/g)	5	0.18	58	15	0.02	24	17	45	—	0.726
	单因子评价	0.33	0.90	0.64	0.40	0.13	0.68	0.43	0.45	0.238	

计算 8 个重金属单因子评价值的标准差发现,土样 3 由于离散程度小,$P_{Nemerow}$ 值略小,但是该点与土样 2 实际的污染程度相近;土样 2 由于单因子最大值 0.95,最小值 0.20,离散度略大,$P_{Nemerow}$ 值居中,但是该点是三者中污染最轻的。

实际情况是 3 个采样点相距不到 3 km,且附近十几千米内的土壤污染程度均为清洁。这个例子表明,对于多因素的度量,内梅罗污染指数由于突出了单因子最大值、区分

度小，影响了对地区污染情况的正确判断。

图 10-1 是传统的内梅罗污染指数评价结果，由于某些采样点被评价为轻度污染，导致在后面的 GIS 插值分析中，该区域产生了一小片轻度污染区，造成污染程度被扩大了，影响对当地污染情况的正确判断。

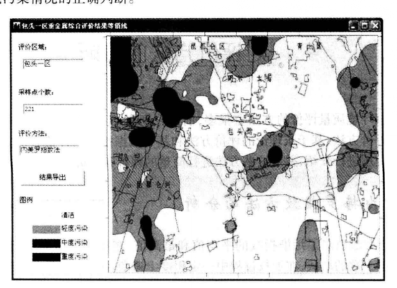
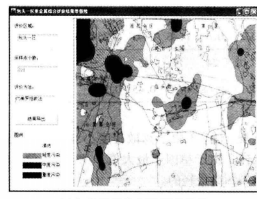


图 10-1　内梅罗污染指数评价的结果

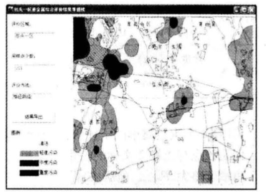
对于多因素的度量，内梅罗污染指数由于突出了单因子最大值，区分度小，甚至有误。此外，不同的重金属应该按照毒性的不同，对土壤危害程度的不同赋给权重值，在内梅罗污染指数中也没有体现。

10.2　对神经网络方法的分析

图 10-2(a)是传统内梅罗污染指数评价的结果，图 10-2(b)是按照前述方法建立的 BP 神经网络模型的评价结果。



(a)内梅罗污染指数评价结果



(b)神经网络评价结果

图 10-2　BP 与内梅罗的对比

两者反映的该地区的污染水平及趋势一致。一些由于内梅罗污染指数评价污染程度过高的地方，按照 Kriging 插值会虚拟出很多中度污染的地方。按照 BP 神经网络模型评价，则有较符合实际的等级划分。例如，在包头市区内存在 1 个 Pb 的点污染区，根据单因子评价的结果，该点的铅污染数为中度污染，该点的综合评价也为中度污染，但是根据内梅罗污染指数法的结果绘制的等值线显示，该区域存在 1 个重度污染点。在包头的农村土壤基本上为清洁，实地调查也不存在污染源，但是内梅罗污染指数法评价结果显示该区域包含 4 个轻度污染点。

由此可知，神经网络的评价结果更符合当地污染的客观情况。

总的来说，可以利用 BP 神经网络法对土壤环境质量进行评价，实验得到结论如下：

1）分析对比研究区域内 8 种重金属的内梅罗污染指数综合评价和神经网络综合评价分布图，两者反映的该地区的污染水平及趋势一致，对于某些细节地域的评价神经网络综合评价更为合理。

2）对于典型区域神经网络综合评价训练样本的选择方式进行有必要探讨。经分析得出样本设置应根据该区域的污染情况合理设置，也就是说必须在区域统计特征的基础上，有针对性地利用等差序列设置训练样本。

3）利用 Kriging 插值和 GIS 可视化分析对研究区域进行分析，结果表明昆都伦河沿岸地区沿河存在 1 个带状污染区；市区郊区存在一些点源污染和小片区的轻度污染。

4）采用神经网络方法对土壤重金属污染进行评价时，涉及到样本和评价标准的选取，对不同区域的污染情况，必须根据样本的统计特性，有针对性地设计训练样本集。

10.3　对灰色关联方法的分析

为了验证灰色关联分析法对土壤重金属污染评价结果的科学性，下面以具体数据和传统的内梅罗综合指数法进行对比，相关数据见表 10-2 所示。

表 10-2　灰色关联土壤重金属污染评价的示例

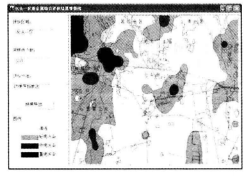
采样点	$P_{Nemerow}$	污染程度		As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
BT-34	1.690	轻度	含量/(μg/g)	6	0.69	54	20	0.3	25	24	6
			单因子评价	0.4	2.23	0.60	0.57	1.18	0.63	0.69	0.7
BT-87	2.528	中度	含量/(μg/g)	5	0.6	57	24	0.06	29	233	572
			单因子评价	0.33	2	0.63	0.69	0.39	0.73	1.63	3.36
BT-111	0.857	轻度	含量/(μg/g)	7	0.2	57	24	0.14	25	20	57
			单因子评价	0.47	1	0.63	0.68	0.93	0.63	0.57	0.57
BT-177	0.563	清洁	含量/(μg/g)	6	0.13	54	13	0.02	24	19	38

续表 10-2

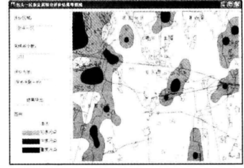
采样点	$P_{Nemerow}$	污染程度		As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
			单因子评价	0.40	0.65	0.60	0.37	0.13	0.60	0.54	0.38
BT-178	0.870	轻度	含量/(μg/g)	5	0.24	49	16	0.04	19	21	45
			单因子评价	0.33	1.10	0.54	0.46	0.29	0.48	0.60	0.60
BT-179	0.729	清洁	含量/(μg/g)	5	0.18	58	15	0.02	24	17	45
			单因子评价	0.33	0.90	0.64	0.43	0.15	0.60	0.49	0.48

上表土样 BT-34 中 Cd 的含量为 0.69 μg/g,单因子评价达到 2.23,超出了国标中的中度污染指标很多,由于 As 等其他 6 种重金属含量比较低,导致该土样的平均值较低,使得内梅罗污染指数评价为轻度污染。土样 BT-87 中 Zn 的含量高达 572 μg/g,单因子评价为 3.36,远远超出了国标重度污染的指标,但是该土样的内梅罗污染评价指数为 2.527,评价为中度污染。而土样 BT-111 中,除了 Cd 的含量值在轻度污染的临界点上外,其他 7 种重金属含量均在清洁的范围内,评价为清洁较为恰当,但是该点的内梅罗污染评价指数为 0.857,被评价为轻度污染。土样 BT-177、BT-178、BT-179 实际情况是 3 个采样点相距不到 3 km,且附近十几公里内的土壤污染程度均为清洁。但是由于土样 BT-178 中 Cd 超过轻度污染临界值 0.04 μg/g,被评价为轻度污染,导致在后面的空间插值分析中,该区域产生了一小片轻度污染区,造成污染程度被扩大了,影响对当地污染情况的正确判断。通过对比分析,可知采用灰色关联分析法能够比较恰当的评定土样的重金属污染情况。

为了进行比较,图 10-3(a)是传统内梅罗污染指数评价的结果,图 10-3(b)是按照前述灰色关联分析评价的结果。



(a)内梅罗污染指数评价结果



(b)灰关联分析评价结果

图 10-3　灰色关联与内梅罗的对比

根据专题图研究重金属污染的空间分布可以得出结论:受包头一些冶金工业生产的影响,土壤中存在不同程度的 As、Hg、Pb 和 Zn 污染。并且包头土壤环境呈现沿昆都伦河被污染明显的条带特点。参照单一重金属的污染值,分析得到重金属对污染的贡献 Zn>

Pb>As> Cd>Cr>Ni> Hg>Cu；此外，郊区有 As、Cd、Pb 和 Zn 的点源污染。从专题图可以看出土壤环境质量在昆都伦河沿岸和郊区的一些地方有向重度污染发展的趋势。实地调研发现导致污染的原因，主要是昆都伦河沿岸地区一些工业企业将工业废水直接排入河中，致以沿岸土壤中重金属元素的含量严重超标；郊区土壤污染主要是尾矿影响。如果不采取有效的治理措施，土壤污染程度的发展必将影响农作物生长、食品安全、人民的生活质量和身体健康。

两种方法所得专题图，反映出研究区域的污染水平及空间分布趋势一致，即包头工业区内土壤污染严重，并且沿昆都伦河成条带分布，市区郊区存在一些点污染。但是在细节上有差异，如在包头的农村土壤基本上为清洁的，实地调查也不存在什么污染源，但是内梅罗污染指数法评价结果该区域包含四个轻度污染点，灰色关联分析法的评价结果更符合农村污染的客观情况；郊区中零散分布着一些点污染区，经实地勘察为一些小型的冶炼厂，灰色关联分析对这些点污染的污染等级评定更为准确；在工业区中，包兰线和昆都伦河交叉的采样点 BT-73，当时采样时采了两个土样，一是按照四分法进行采样，该采样点内梅罗污染指数为 1.23，为中度污染；另外只在一堆尾矿附近采土样，该土样的铅含量达 3 075 μg/g，内梅罗污染评价指数为 11.5，为重度污染。将这两个土样都标绘在专题图上时，Kriging 插值会虚拟出很多重度污染的地方；而灰色关联分析方法由于先将单因子污染指数进行归一化处理，可以避免空间插值环节可能产生的虚拟出重度污染的情况。

研究区域包头及郊区的土壤重金属污染呈现沿昆都伦河被污染明显的条带特点，污染程度最严重的出现在工业区；并且重金属对污染的贡献 Zn>Pb>As>Cd>Cr>Ni>Hg>Cu。

总之，土壤中的污染因素的多样性及复杂性，决定了要直接分析重金属指标与土壤环境质量的关系很困难。目前应用最广泛的回归分析法和内梅罗指数法都存在缺陷，如要求大量数据、计算量大、精度不高甚至可能出现反常情况等。灰色关联分析方法根据因素间发展态势的相似或相异程度来衡量因素的关联程度，因此用于评价土壤重金属污染可以获得更为真实的评价结果。

10.4 对数据挖掘方法的分析

土壤作为一种 GIS 空间对象，既具有非空间的属性特征，又具有与位置相关的空间特征。如果不考虑其空间特征，单一使用非空间的属性特征来进行聚类，聚类结果会与实际情况产生较大的差异，不能全面的反应对象的内在联系。考虑空间格局的谱聚类算法考虑了不同对象的空间相邻性，从而使聚类方案可反映不同类别对象的空间分布状态。因此更适合于对空间对象的聚类分析。

由于数据挖掘方法单纯从数据的角度出发，没有考虑土壤污染的物理过程，在研究中只是作为尝试性的分析，以土壤重金属形态数据作为研究对象，以求揭示重金属形态对土壤环境质量的影响及决定性作用。

第 11 章

典型应用

本章介绍两个典型的应用示例，分别是对包头和苏州研究区的土壤重金属污染的研究。

11.1 应用实例一

土壤重金属污染评价及安全预警系统的设计，为实现一种新的土壤重金属污染评价提供了基础平台，为土壤环境质量的预测预警提供了基础和参考。研究开发的平台可以对大气、植物、水体等环境要素实现监测，为城市环境综合质量监控提供基础数据。下面以两个应用为例，示范平台的主要功能。

11.1.1 研究区概述

包头市区中心的经纬度为北纬 40.39°，东经 109.49°。作为中国能源重化工基地的中心城市，包头及其城市区域的重金属污染状况具有典型性和代表性。包头钢铁厂是典型的污染源，它是“一五”期间建设的老企业，初期环保设计不足，大部分主体设施没有配套建设环保设施，部分主体设施虽然配套了环保设施，但随着包钢产能的不断扩大，环保设施的处理能力明显不足。

研究区域是一个矩形的地理范围，如图 11-1 所示，面积 221 km^2，涵盖了包头的工业区、市区、郊区、农村等城市功能区。



图 11-1　包头研究区域采样示意图(源自 Google 地图)

11.1.2 土壤采样及分析

野外实验在该采样区采集了 222 个土壤表层样品，还有一些土壤剖面样品、大气样品、植物样品、水样等。下面重点介绍土壤样品的采集及分析。

采样过程严格依据《土壤环境监测技术规范》进行操作。采样前，准备了包头地区 1∶5万地图、行政区划图、土壤类型分布图、重金属元素分布图、地球化学图、城市点源污染分布图、监测点位分布图等资料。采样时，携带的采样器具有：木铲、刮刀、广泛 pH 值试纸、GPS、罗盘、照相机、卷尺、布制样品袋、样品箱、标签纸、采样记录表、铅笔、资料夹等。乘坐的交通工具为越野车。

为了研究不同城市功能区的表层土壤重金属污染情况，经过现场踏勘，首先划分包头市及郊区的城市功能区，主要为工业区、市区、郊区、农村，如图 11-2 所示。其中各个功能区域的大致面积为：工业区 35 km^2，市区 36 km^2，郊区 65 km^2，农村 85 km^2。

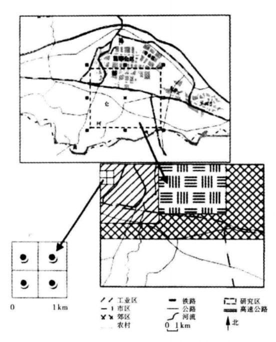


图 11-2　包头研究区采样布点示意

布点方法为系统随机布点法。将研究区域划分为 1 km^2 网格，以单个网格为一个采样布点基本单位，网格内采用四分法布点。选点的原则为在 1 km^2 的范围内，尽可能选取具有相同使用功能的土壤类型。

土壤样品采集的具体方法是：先拨开覆盖在表层土壤最上层的树枝、碎石和杂草，用木铲垂直于地表，挖出 0～20 cm 的土壤剖面表层，取厚度大致相同的土壤薄片，观察土壤特征，除去大颗粒杂质，取约 1 kg 左右，装入采样布袋中，扎口、编号。采集过程得到土壤样品包括：工业区 140 件、市区 144 件、郊区 260 件、农村 340 件。采集的每个土壤剖面

样共有 7 个剖面层，距离地表的深度分布为 20 cm、40 cm、60 cm、80 cm、100 cm、150 cm、200 cm。

样品进行装样和编号时，必须同时写两个标签，装样袋内外各一个标签，注明采样地点、层次厚度、天气、时间、采集人等，经火车运输至北京。

野外采集来的土壤样品放在通风处，拣去根系及残枝落叶，摊开风干，避免暴晒。等待自然风干后，用塑料锤将土壤样品捣碎，将一个采样网格内 4 个采样点采集的土壤样品进行混合，每个样品中取出大致相同质量的土壤样品，均匀混合，组合成一个混合样，重新编号。共获得 221 件土壤混合样品。

本研究将混合样送到国家地质实验测试中心进行分析测定，该单位实验室通过国家实验室认可（证书编号为 CNAS L1299），具备计量认证资质（2007000583G）。

由于重点分析测定土壤样品中的重金属含量，所以在样品采样、风干、研磨、过筛、运输、贮存等环节都要特别注意，不能接触金属器具，以防土壤样品被污染。在实验室中，将样品平铺在制样板上，用木棍或塑料棍碾压，并将植物残体、小石块等杂质剔除干净，细小已断的植物须根，可采用静电吸附的方法清除。若在土壤中有铁锰结核、石灰结核、铁子或半风化体，不能用木棍碾碎，应细心拣出称量保存。最后将土壤混合样再次研磨，过孔径为 880 μm（200 目）尼龙网筛，送检测室等待测试。

经过国家地质实验测试中心专业人员工作，分析测量了土壤混合样品中的 8 种重金属含量。As 和 Hg 采用原子荧光光谱法测定，Cd、Cr、Cu、Ni、Pb、Zn 采用等离子体质谱法测定。并且，还对各种重金属分析了水溶态、离子交换态、碳酸盐结合态、腐殖质结合态、铁锰氧化物结合态、有机结合态、残渣态等形态数据。样品测试过程中，每十个样品以密码形式加国家一级标准物质（GSS 土壤系列）进行质量监控，以确保分析数据的准确可靠。

11.1.3 实现功能

在展开土壤环境质量评价之前，首先建立空间数据库和属性数据库，整合多源数据。建立的电子地图包括公路、铁路、湖泊、河流、居民点、土壤剖面采样点、大气采样点、土壤采样点、网格等图层，叠加后的研究区域电子地图如图 11-3 所示。

在土壤重金属污染评价及安全预警系统平台的支撑下，包头研究区的工作空间如图 11-4所示。中间最大的部分是地图操作窗口，系统的功能以菜单和对话框的形式组织。下面将对具体功能逐一说明。

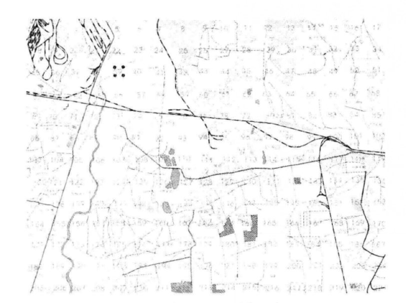


图 11-3　包头研究区的电子地图

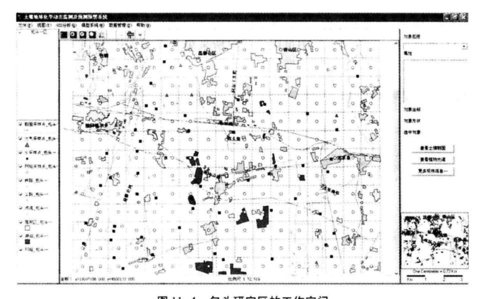


图 11-4　包头研究区的工作空间

图中的小圆圈代表网格采样点，表示该处采集的有土壤样品和植物样品，点击一个采样点后，选择“专题分析”子菜单中的“植物光谱”，执行光谱分析的结果如图 11-5 所示，该图显示的是包头一区 BT-18 采样点的油菜植物叶片光谱分析结果。

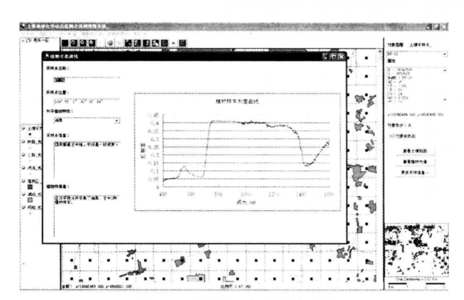


图 11-5　包头植物叶片光谱分析结果

专题图功能为单一重金属含量区域整体情况研究提供了方便，专题图分析针对某一个工作区间，选择“专题图分析”子菜单，执行土壤重金属含量数据的专题图分析，分析结果如图 11-6 所示。下拉列表框中可以选择对所采集土壤样本的 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属含量数据的专题图。

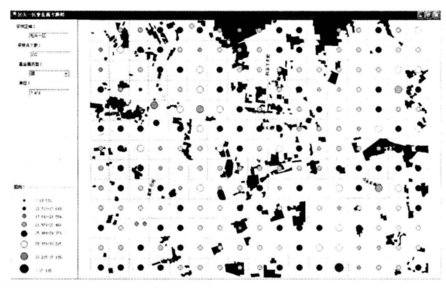


图 11-6　包头一区土壤重金属 Ni 专题分析结果

土壤重金属污染评价及安全预警系统还可以对区域单一重金属的污染情况进行定性分析。例如采样国标单因子评价方法，选择“区域评价模型”子菜单，执行采样区域单

一重金属评价分析,分析结果如图 11-7 所示。根据土壤中的 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属的污染情况,划分为清洁、轻度污染、中度污染、重度污染 4 个等级,将整个区域的污染等级绘制成等值线,如图 11-7 所示。

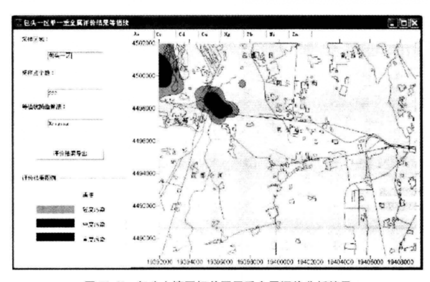


图 11-7 包头土壤国标单因子重金属评价分析结果

选择"国家标准综合评价模型"子菜单,执行土壤重金属国家标准综合评价模型评价分析,分析结果如图 11-8 所示。

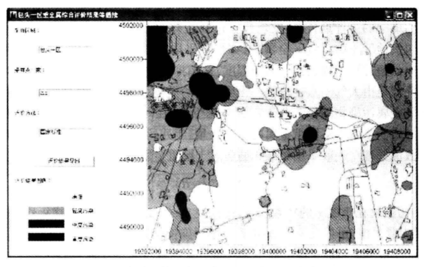


图 11-8 包头土壤国标综合因子重金属评价分析结果

选择"区域评价模型"子菜单,执行土壤重金属神经网络综合评价模型评价分析,分析结果如图 11-9 所示。

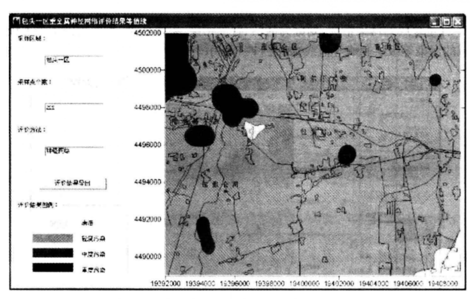


图 11-9　包头土壤神经网络重金属评价分析结果

研究结果表明，受包钢生产的影响，研究区域土壤中存在不同程度的 As、Cd、Zn、Cu 等重金属污染。从预警结果可以看出，土壤污染中的重金属在昆都伦河沿岸的采样点和郊区的一些地方正在向污染超标的方向演化。借助本系统，确定了包头地区的污染主要是昆都伦河沿岸地区一些工业企业将工业废水直接排入河中，导致沿岸土壤中重金属元素的含量严重超标，郊区主要是尾矿影响，分析结果与实地调查的真实情况一致。综合考虑土壤污染中的多种重金属，发现包头市及近郊的土壤质量不容乐观，根据预测结果可以给出包头地区土壤质量安全预警，如果不采取有效的治理措施，预计该地区将在未来较短时间内超过预警级别达到污染程度，这必将影响人民的生活质量和身体健康。

城市环境是由大气—水—植物—土壤组成的循环系统，是一个微妙的平衡系统。假定土壤中重金属污染物随着工业生产和人类生活，在时间维度上积累，并且其值可以用通量计算。由于土壤中不同重金属的年输入量、年残留率均不同，通过建立矩阵，可以方便运算和数值处理。一般采用年累积通量对土壤重金属含量进行数值预测：

$$W=K(B+R)$$

其中，W 为污染物在土壤中的年累积量，单位是 mg/kg；K 为土壤污染物年残留率，是一个百分比；B 为背景值，单位是 mg/kg；R 为土壤污染物年输入量，单位是 mg/kg。

K 值可以通过实验室土壤的吸附和解吸试验计算得到。计算公式为：

$$K = s/(s+s')$$

其中 s 为土壤中重金属的吸附量，单位是 mg/kg；s' 为土壤中重金属的解吸量，单位是 mg/kg。吸附量反映了土壤被污染的程度，解吸量反映了土壤的自净能力，K 值综合反映了土壤被污染的程度。

为了预测土壤环境质量的变化趋势，基于年累积量的预测预警模型，可以实现数据模拟预测、空间模拟预测、专题模拟预测等多个方面的分别处理。

年累积通量数值预测模型基于的假定是：土壤中重金属污染物随着工业生产和人类生活，在时间维度上积累，并且其值可以用通量计算。该模型考虑到土壤的净化能力，为污染物累积提供了数值预测的方法。

在对土壤重金属污染年累积通量数值预测的基础上，可以展开基于 GIS 的环境质量预测预警。选择国标综合因子模型来对预测数据进行评价，并绘制等值线图，分析 5 年、10 年、20 年土壤环境质量的时空演化规律。该模型假定土壤维持现有特性，并且不施加环境保护措施。根据年累积通量预测得到的重金属污染新数据，依据国家环境质量标准（GB 15618—1995）对 n 年后某区域中土壤重金属污染情况进行评价，从而实现短期、中期、长期的空间分析预警，这是快速准确地找到污染源并实现土壤重金属污染的安全预警的基础。

使用该模型进行土壤重金属环境质量预测预警的流程，如图 11-10 所示。

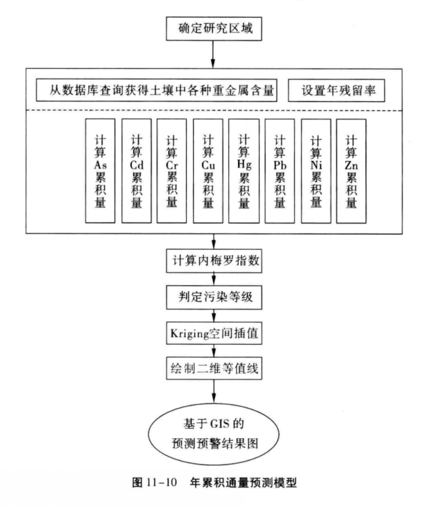


图 11-10　年累积通量预测模型

选择“预测分析模型”子菜单，执行采样区域 5 年、10 年、15 年、20 年后土壤重金属的

污染情况分析，按照国家标准，将污染情况划分为四个等级，并绘制成等值线，分析结果如图 11-11 所示。

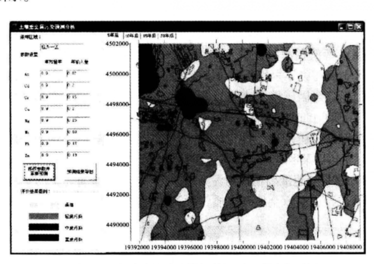


图 11-11　包头土壤重金属污染预测预警分析结果

如果在预测分析模型参数设置区域修改了模型的某个参数，单击“修改参数并重新预测”按钮，然后再次选择“预测分析模型”子菜单，执行新的预测分析。

环境质量预测预警模型是基于年累积通量数值预测模型的数据预测结果，参照国标，使用国标综合因子法来评价土壤环境质量。模型设计土壤中各种重金属的残留率由实验室土壤实验获得，模型设计中其值是可调节的，这在一定程度上也说明增加人工手段可以干涉重金属在土壤中的累积，能够影响环境质量。

11.2　应用实例二

苏州位于江苏省东南部，东临上海，南接浙江，西抱太湖，北依长江。市区中心地理坐标为北纬 31°19′，东经 120°37′。随着苏州城市化进程的加速，大量工业园区的建设，苏州城市环境污染问题日益严重。

11.2.1　苏州研究区概述

苏州研究区是一个矩形的地理范围。研究人员在该采样区采集了土壤剖面样、土壤千米网格样、植物样品等。经过实验室分析测试，将预处理的空间数据和属性数据规范标准化后录入数据库。

11.2.2　实现功能

在平台的支撑下，苏州研究区的工作空间如图 11-12 所示。

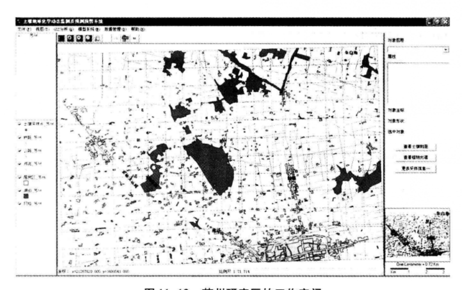


图 11-12　苏州研究区的工作空间

土壤剖面采样点的深度间隔为 20 cm 一个剖面层，所分析的重金属有 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属，各种重金属分别以曲线拟合出剖面线。选择“专题分析”子菜单中的“土壤重金属剖面”，执行分析的结果如图 11-13 所示。

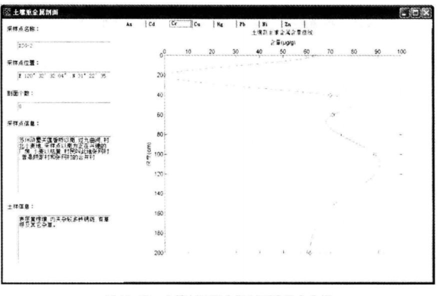


图 11-13　土壤剖面重金属剖面线拟合分析

土壤采样点所采集土样的重金属形态包括水溶态、离子交换态、碳酸盐结合态、腐殖质结合态、铁锰氧化物结合态、有机结合态和残渣态，每种重金属即 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属各有 7 种形态，分别以饼图显示。“all”选项卡中将所有重金属形

态以百分比堆积柱状图输出。选择“专题分析”子菜单中的“土壤重金属形态”，执行分析结果如图 11-14 所示。

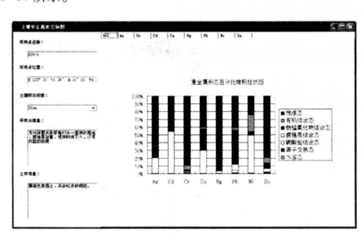


图 11-14　土壤重金属形态百分比堆积分析

选择“专题分析”子菜单中的“植物重金属含量”，执行分析的结果如图 11-15 所示。采样点所采集的所有植物种类以列表形式显示，在植物种类下拉列表框中选择一种植物，则该植物的采样点、采样部位等信息即在左边对应文本中显示。所分析的重金属有 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属，植物各部位重金属含量的柱状在右侧图显示。

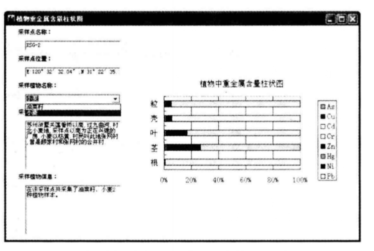


图 11-15　植物各部分重金属含量数据分析

苏州作为城市化进程较快的典型代表城市，其城市环境重金属污染主要来源于工业污染。传统城市环境监测的研究区域常局限在县、市单元或其他行政单元，研究对象单一，指标的选取偏向自然，图件编制和数值量算上多采用手工操作，而且研究结果多以文字形式表现出来。本研究所开发的平台，能够将许多掩埋在文字背后的深层次信息挖掘出来，增强了统计分析、可视化分析及决策支持等功能。

第 12 章

系统使用方法

本章详细介绍系统的使用方法，大体分 5 个功能模块：工作空间、GIS 空间分析、专题分析、评价及预测模型、数据管理。

12.1 工作空间的使用

系统以工作空间的形式组织、管理研究区域的图层数据。打开文件菜单包括工作空间（即典型区域）、图层控制及退出整个系统等选项。

12.1.1 工作空间

根据研究需要，系统针对 6 个典型区域实现统一管理，分别是包头一区、包头二区、苏州、崇明一区、崇明二区和新江湾，如图 12-1 所示。

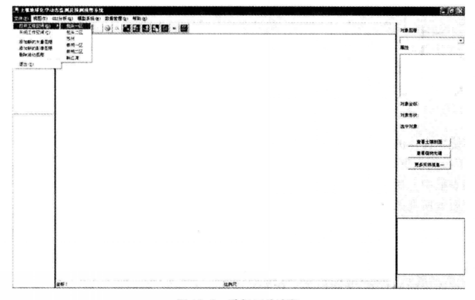


图 12-1 选择工作空间

单击“文件”菜单下的“打开工作空间”子菜单，指向次级菜单，选中需要的工作空间，单击后可以在主窗口区域加载该工作空间的所有图层。以包头一区为例，界面如图12-2所示。

系统主窗体的中间部分为地图显示区域，将工作空间的多个图层叠加来展示研究区域的主要地理要素。

系统主窗体的左侧上半部分为工作空间的所有图层的树状列表。选中某一个树节点，单击鼠标右键，会显示文件操作的快捷菜单，包括添加新的矢量图层、删除活动图层、打开工作空间、关闭工作空间等选项。

系统主窗体的左侧下半部分为工作空间的所有图层的拣选列表，单击图层名称前面的拣选框。对号表示选中，并在地图显示区域显示该图层；没有对号表示不选，在地图显示区域将该图层隐藏。

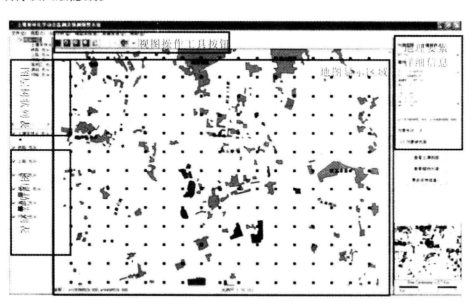


图12-2　加载包头一区工作空间后显示的所有图层

系统主窗体的右侧上半部分为地理要素的详细信息，包括“对象图层”文本框、“地理要素”下拉列表框和“属性”文本框。在地图显示区域选中的所有地理要素，将逐一列在下拉列表框中。用鼠标在列表框中选中一个地理要素，可以在“对象图层”文本框中显示该地理要素所在的图层，在“属性”文本框中显示该地理要素的所有属性。

系统主窗体的右侧下半部分为工作空间的鹰眼视图，与地图显示区域的放大、缩小、平移等联动更新。

系统主窗体的底部为地图比例尺和公里网坐标信息。

对工作空间的打开、关闭以及工作空间中图层的增加、删除有两种方法：使用文件菜单，或者使用右键快捷菜单。

如果需要同时操作多个工作空间，必须先选中图层树状列表的根节点。具体方法

为：鼠标指向系统主窗体左侧上半部分的工作空间所有图层树状列表，选中某一个根节点（即某一个工作空间），在“文件”菜单中或右键快捷菜单中，选择并执行“打开工作空间”或“关闭工作空间”。

对工作空间中图层的操作方法为：鼠标指向系统主窗体左侧上半部分的工作空间所有图层树状列表，选中工作空间树的某一个叶节点（即某一个要操作的图层），移动鼠标到文件菜单，单击打开后选择需要的子菜单；或者选中某一个叶节点，单击鼠标右键，在显示的快捷菜单中，选择并执行“添加新的矢量图层”、“删除活动图层”等。

如果需要调整地图显示区域中图层的叠加层次，必须先选中图层拣选列表的某一个图层。具体方法为：鼠标指向系统主窗体左侧下半部分的工作空间所有图层拣选列表，选中某一个要操作的图层，拖动鼠标以改变图层的叠加顺序。图层拣选列表的第一个图层为活动图层，在地图显示区域的最顶层显示。

单击文件菜单中的“退出”子菜单，可以关闭系统。该功能对应的快捷键为“Ctrl+X”。

12.1.2 视图

视图菜单包括：放大、缩小、全图、平移、查找、选择等常用的地图操作。

要将地图显示区域中感兴趣区域放大，选择视图菜单中的“放大”子菜单，或者单击地图常用操作中的“放大”工具按钮，移动鼠标到地图显示区域，这时鼠标的形状变成放大镜，单击一次则以点击处为中心放大一倍。

要将地图显示区域中感兴趣区域缩小，选择视图菜单中的“缩小”子菜单，或者单击地图常用操作中的“缩小”工具按钮，移动鼠标到地图显示区域，这时鼠标的形状变成缩小镜，单击一次则以点击处为中心缩小一倍。

要将地图显示区域进行平移，选择视图菜单中的“全图”子菜单，或者单击地图常用操作中的“全图”工具按钮，移动鼠标到地图显示区域，这时鼠标的形状变成全图镜，单击一次则恢复到地图全部显示。

要将放大或者缩小后的地图显示区域恢复到全部显示的状态，选择视图菜单中的“平移”子菜单，或者单击地图常用操作中的“平移”工具按钮，移动鼠标到地图显示区域，这时鼠标的形状变成手形，拖动鼠标可以移动地图显示区域中所加载的图层。

12.1.3 按条件查询和框选

如果知道所要查找地理要素的部分条件，可以选择该地理要素所在的图层，然后输入查询条件，点击查询按钮，即可找到该地理要素。

如果仅知道所有查找地理要素的大致位置，可以使用框选。选择视图菜单中的“框选”子菜单，或者单击地图常用操作中的“框选”工具按钮，可以打开地理要素框选功能。在地图显示区域中拖动鼠标，框选一片感兴趣的区域，则找到的该区域中所有的地理要素会闪烁变色三次，并在系统主窗体的右侧上半部分显示所有地理要素的详细属性信息。

12.2 GIS 空间分析的使用

“GIS 分析”菜单包括:空间分析、专题分析、统计分析、专题图分析等常用的基于电子地图的空间要素分析和属性数据分析。本系统的空间分析包括采样点查询和距离量算两个功能。

12.2.1 采样点查询

选择“空间分析”子菜单中的“采样点查询”,或者单击地图常用操作中的“查找”工具按钮,可以打开地理要素查找功能,如图 12-3 所示。

在图层下拉列表框中选择要查找的图层,在查找条件文本框中填写查找的条件,单击查找按钮执行查找功能。查找到的地理要素会闪烁变色三次,并在系统主窗体的右侧上半部分显示该地理要素的详细属性信息。

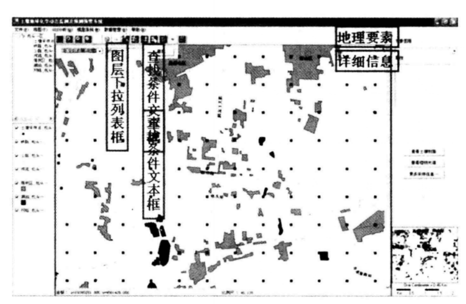


图 12-3 指定条件查找地理要素

12.2.2 距离量算

选择“空间分析”子菜单中的“距离量算”,或者单击地图常用操作中的“量算”工具按钮,可以执行地图上任意路线量算的功能,如图 12-4 所示。

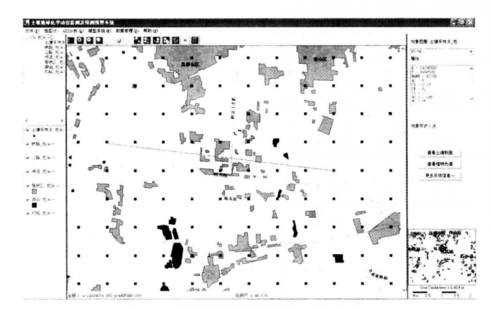


图 12-4　鼠标选择距离量算的起点、折点和终点

12.3　专题分析的使用

专题分析主要针对土壤和植物，包括：土壤重金属形态、剖面、等值线和植物光谱、重金属含量等五个功能。

要执行专题分析功能，必须先选择一个土壤采样点，在该采样点采集的土样的重金属检测数据则会显示在系统主窗体的右侧上半部分，即将 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等土壤重金属含量作为了该采样点的详细属性信息。然后可以分别执行“专题分析”子菜单下的任何一个次级菜单。

12.3.1　土壤重金属形态

执行“专题分析”子菜单中的“土壤重金属形态”分析功能，需要先选择一个土壤采样点，如图 12-5 所示。执行分析功能的结果如图 12-6 所示。

土壤采样点所采集的土样的重金属形态都包括水溶态、离子交换态、碳酸盐结合态、腐殖质结合态、铁锰氧化物结合态、有机结合态和残渣态，即 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属各有 7 种形态以饼图分别显示，如图 12-6 所示。all 选项卡中将所有重金属形态以百分比堆积柱状图显示，如图 12-7 所示。

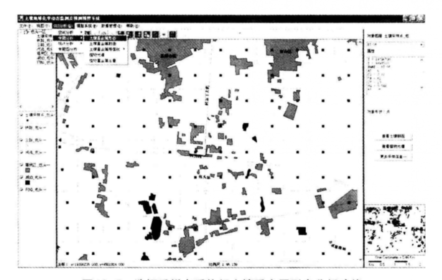


图 12-5　选择采样点后执行土壤重金属形态分析功能



图 12-6　土壤重金属形态分析饼图

注意,剖面采样点由于剖面层数较多,画图可能较慢,在土壤样本深度下拉列表框中选择不同的样本深度可以查看不同的剖面层。如果是网格采样点,土壤样本深度下拉列表框显示为 0 cm,如图 12-7 所示。

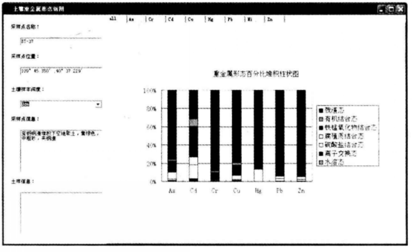


图 12-7　土壤重金属形态分析结果

12.3.2　土壤重金属剖面

选择“专题分析”子菜单中的“土壤重金属剖面”分析功能，先选择一个土壤采样点，如图 12-8 所示，再执行分析功能。

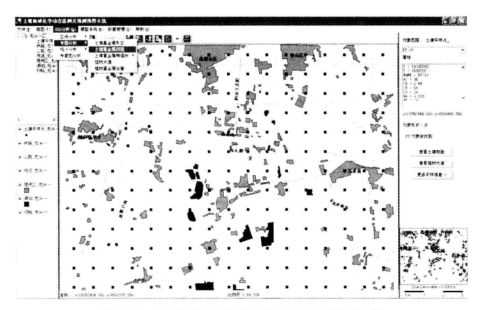


图 12-8　选择采样点后执行土壤重金属剖面分析功能

土壤剖面采样点的深度为 20 cm 一个剖面层，所分析的重金属有 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属，各种重金属分别以曲线拟合出剖面线，如图 12-9 所示。

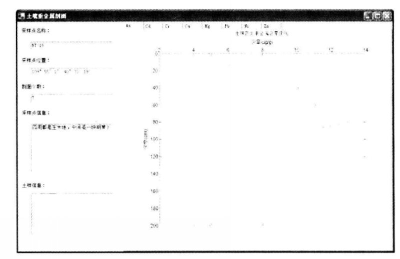


图 12-9　土壤重金属剖面分析结果

12.3.3 土壤重金属等值线

土壤重金属等值线分析的结果有两种展示方式，分别是2D效果显示和3D效果显示。

选择"专题分析"子菜单中的"土壤重金属等值线"，执行2D效果显示的结果如图12-10所示。

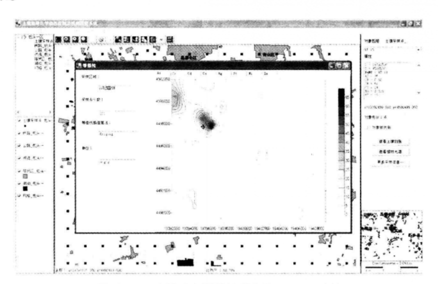


图12-10 土壤重金属等值线分析2D显示功能

以采样区域的所有采样点为基础数据，所分析的重金属有As、Cd、Cr、Cu、Hg、Ni、Pb、Zn等8种重金属，各种重金属分别拟合出区域2D效果的等值线。

选择"专题分析"子菜单中的"土壤重金属等值线"，执行3D效果显示如图12-11所示。

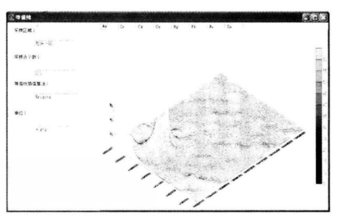


图12-11 土壤重金属等值线分析3D显示结果

以采样区域的所有采样点为基础数据，所分析的重金属有 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属，各种重金属分别拟合出区域 3D 效果的等值线。

12.3.4 植物光谱

选择“专题分析”子菜单中的“植物光谱”，执行光谱分析的结果如图 12-12 所示。

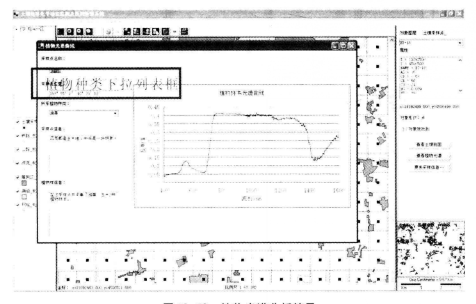


图 12-12　植物光谱分析结果

采样点所采集的所有植物种类以列表形式显示在窗体左侧，在植物种类下拉列表框中选择一种植物，该植物的光谱曲线在右侧显示。

12.3.5 植物重金属含量

选择“专题分析”子菜单中的“植物重金属含量”，执行分析的结果如图 12-13 所示。

采样点所采集的所有植物种类以列表形式显示在窗体左侧，在植物种类下拉列表框中选择一种植物，该植物的根、茎、叶、壳、粒等部位重金属含量分别在实验室测得。所分析的重金属有 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属，植物各部位重金属含量的柱状图显示在右侧。

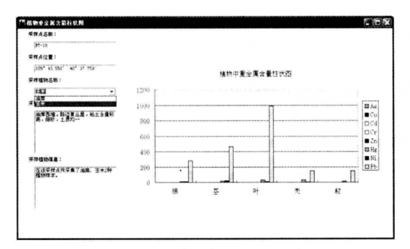


图 12-13　**植物重金属含量分析结果**

12.3.6　水样中重金属含量

选择"专题分析"子菜单中的"水样中重金属含量"，执行分析的结果如图 12-14 所示。

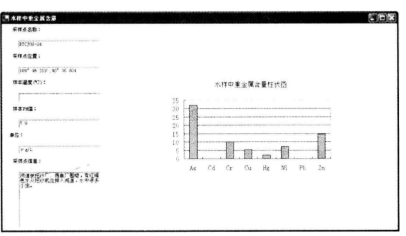


图 12-14　**水样重金属含量分析**

12.3.7　大气样中重金属含量

选择"专题分析"子菜单中的"大气样中重金属含量"，执行分析的结果如图 12-15 所示。

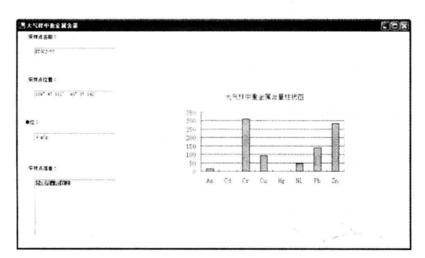


图 12-15 大气重金属含量分析

12.3.8 统计分析

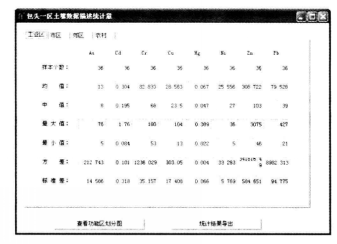
统计分析主要针对数据库中的植物、大气、土壤等样本数据进行基本的描述统计量分析。

选择"统计分析"子菜单中的"水样数据"，执行水样数据的统计分析。

选择"统计分析"子菜单中的"大气数据"，执行大气数据的统计分析。

选择"统计分析"子菜单中的"土壤数据"，执行土壤数据的统计分析。

以包头一区土壤数据的统计分析结果为例，如图所示 12-16 所示。基本的描述统计量包括：样本个数、均值、中值、最大值、最小值、方差、标准差。将包头一区按照功能区划分为工业区、市区、郊区和农村，分别对各功能区所采集土壤数据中 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属的样本进行统计。



	As	Cd	Cr	Cu	Hg	Ni	Zn	Pb
样本个数：	36	36	36	36	36	36	36	36
均 值：	13	0.304	82.833	28.583	0.067	25.556	308.722	79.528
中 值：	8	0.195	68	23.5	0.047	27	103	39
最大值：	76	1.76	180	104	0.389	36	3075	427
最小值：	5	0.084	53	13	0.022	5	46	21
方 差：	212.743	0.101	1236.029	303.05	0.004	33.263	341516.49	8982.313
标准差：	14.586	0.318	35.157	17.408	0.066	5.769	584.651	94.775

图 12-16 包头一区土壤数据统计分析结果

单击“查看功能区划分图”按钮，可以查看包头一区城市的四个功能区：工业区、市区、郊区和农村。单击“统计结果导出”按钮，可以将土壤数据的统计分析结果导出到 Excel 中。

12.3.9 专题图分析

专题图分析针对某一个典型区域，对所采集土壤样本的 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属含量数据绘制专题图。选择“专题图分析”子菜单，执行土壤重金属含量数据的专题图分析，如图 12-17 所示。

在“重金属种类下拉列表框”中选择一种重金属，则土壤中该重金属含量数据的专题图显示在右侧。

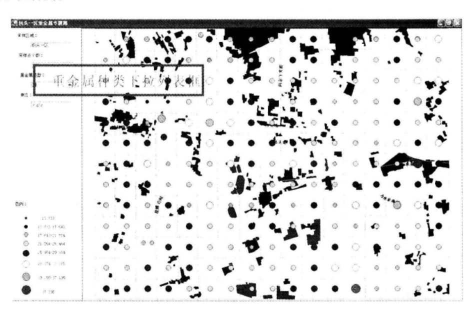


图 12-17　包头一区土壤数据统计分析结果导出

12.4 评价及预测模型的使用

本系统实现了典型区域土壤状况监测的评价和预测，主要包括单点评价和区域评价，以及预测分析模型。

单点评价模型：根据某一个采样点的土壤数据，分析评价该采样点的土壤重金属污染等级指数。

区域评价模型：对整个工作空间的土壤数据分析评价，分析评价整个区域的土壤环境污染等级指数，以等值线图显示。

预测分析模型：根据整个工作空间的土壤数据，对整个区域的土壤环境质量分析预测。

“模型系统”菜单中包括单点评价模型、区域评价模型和预测分析模型。评价模型又分为 3 种：单一重金属因素评价、神经网络综合评价和国家标准综合评价。

12.4.1 单点评价模型

单点评价模型有3个:单一重金属评价模型、神经网络综合评价模型和国家标准综合评价模型。注意,单点评价模型的前提是选择了某一个网格采样点。

选择"单点评价模型"子菜单,执行土壤重金属单一重金属评价分析。

以包头一区某一采样点为例,该采样点土壤数据中As、Cd、Cr、Cu、Hg、Ni、Pb、Zn等八种重金属的单因子评价结果污染指数以柱状图显示,如图12-18所示。单击"评价结果导出"按钮,可以将该采样点重金属污染的评价结果导出到Excel中,如图12-19所示。

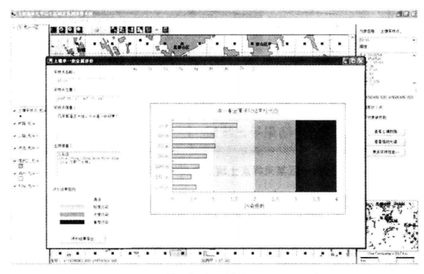


图12-18 包头一区某采样点土壤单一重金属评价分析结果

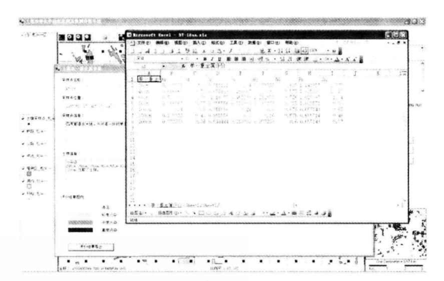


图12-19 包头一区某采样点土壤单一重金属评价分析结果导出

选择“神经网络综合评价模型”子菜单，执行土壤重金属神经网络综合评价模型评价分析。分析结果如图 12-20 所示。

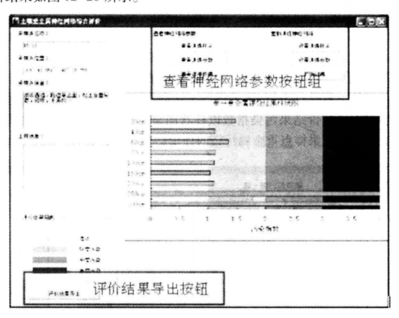


图 12-20　包头一区某采样点土壤重金属神经网络综合评价分析结果

单击“评价结果导出”按钮，可以将该采样点土壤重金属神经网络综合评价模型的评价结果导出到 Excel 中，如图 12-21 所示。在“查看神经网络参数”按钮组中，单击“查看训练样本”按钮、“查看训练参数”按钮或“查看连结权值”按钮，可以打开 Excel，详细查看采样点土壤重金属神经网络综合评价模型的具体参数，如图 12-22 所示。

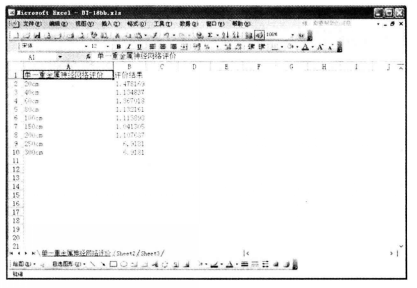


图 12-21　包头一区某采样点土壤重金属神经网络综合评价分析结果导出

	A	B	C	D	E	F	G	H	I	J
1	隐层权值									
2	1.5207	5.0063	0.7619	-0.3036	0.1891	-0.69	-4.2808	1.5285		
3	-0.7089	3.5614	-0.5202	-3.0372	-1.3545	-1.3619	-2.3733	5.273		
4	5.4465	2.4244	5.7589	1.7386	0.1573	-0.8051	6.7554	0.8425		
5	1.4266	-3.1298	0.5235	2.4019	1.08	-0.9288	-0.0406	-14.233		
6	-1.8486	-0.6636	1.7873	2.5262	1.278	-1.3927	-0.2187	-6.2855		
7	1.1468	-4.4996	2.6355	-2.2132	-1.9658	-1.0646	1.8857	5.3686		
8	-1.2906	-2.5371	-0.5798	1.0544	-0.248	-0.8533	1.4615	5.9906		
9	-1.6688	2.7955	-4.3759	-1.8272	-6.9486	-2.3865	1.3219	2.5963		
10	-0.8904	1.4374	1.7537	1.6944	-1.3829	-0.7469	-1.9888	-2.9236		
11	-0.6903	-2.9689	-1.8959	-0.6627	1.9496	0.1072	1.5881	-1.6234		
12	输出层权值									
13	1.4332	-0.427	3.0747	-1.6386	0.3989	0.0841	6.3037	-0.1036	-1.4348	-0.0001
14	隐层偏置值									
15	1.8202	1.4478	-3.6034	-9.1699	-1.9588	1.5599	0.1047	0.3567	-0.0411	-3.8198
16	输出偏置值									
17	3.65									

图 12-22　包头一区某采样点土壤重金属神经网络综合评价模型参数

选择"国家标准综合评价模型"子菜单,执行土壤重金属单一重金属国家标准综合评价模型评价分析,分析结果如图 12-23 所示。

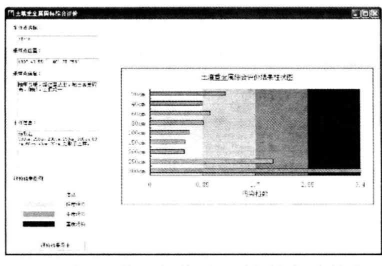


图 12-23　包头一区某采样点土壤重金属国家标准综合评价分析结果

12.4.2　区域评价模型

区域评价模型有 3 个:单一重金属评价模型、神经网络综合评价模型和国家标准综合评价模型。

选择"区域评价模型"子菜单,执行采样区域单一重金属评价分析,分析结果如图 12-24所示。根据 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 等 8 种重金属的污染情况,将每一采样点的土壤污染划分为清洁、轻度污染、中度污染、重度污染等 4 个等级,再将整个区域的污染等级绘制成等值线。单击"评价结果导出"按钮,可以将该采样点重金属污染的评价结果导出到 Excel 中,如图 12-25 所示。

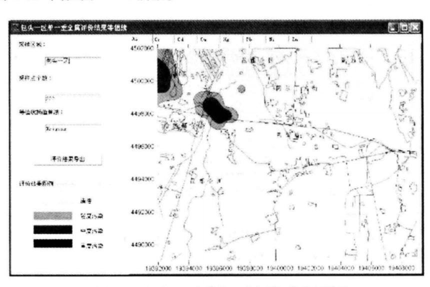


图 12-24 包头一区土壤单一重金属评价分析结果

Microsoft Excel - 包头一区dd.xls

	A	B	C	D	E	F	G	H	I
1	单一重金属评价	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
2	BT-2	0.333333	1.1	0.711111	0.24	0.373333	0.625	1.142857	1.575
3	BT-3	0.666667	0.75	0.688889	0.266667	0.32	0.575	1.025397	1.015
4	BT-4	0.733333	0.75	1.16875	0.453333	1.281176	0.725	1.025397	1.04
5	BT-5	0.866667	0.75	0.755556	0.333333	0.313333	0.725	1.028571	0.83
6	BT-6	0.4	0.5	0.755556	0.28	0.813333	0.675	0.657143	0.65
7	BT-6-A	0.333333	0.42	0.655556	0.2	0.16	0.575	0.685714	0.46
8	BT-7	0.333333	0.38	0.644444	0.253333	0.9	0.6	0.657143	0.57
9	BT-8	0.333333	0.7	0.711111	0.28	0.386667	0.625	0.628571	0.77
10	BT-9	0.4	0.47	0.566667	0.266667	0.62	0.625	0.657143	0.61
11	BT-10	0.333333	0.85	0.7	0.306667	0.353333	0.725	3.966667	0.7
12	BT-11	0.4	0.6	0.711111	0.28	0.373333	0.65	0.685714	0.57
13	BT-12	0.333333	0.65	0.688889	0.24	0.153333	0.675	1	0.44
14	BT-13	0.4	0.41	0.733333	0.213333	0.16	0.55	0.542857	0.41
15	BT-14	0.4	0.4	0.6	0.2	0.153333	0.55	0.485714	0.45
16	BT-15	0.466667	0.6	0.855556	0.293333	0.24	0.675	0.628571	0.57
17	BT-16	0.533333	0.47	0.666667	0.24	0.18	0.625	0.542857	0.48
18	BT-17	0.466667	0.35	0.733333	0.413333	0.126667	0.75	0.571429	0.54
19	BT-18	2.733333	2.2	0.755556	0.32	0.26	0.475	1.38254	6.52
20	BT-19	0.4	0.7	0.7	0.346667	0.26	0.75	1.019048	1.035
21	BT-20	0.466667	1.325	0.844444	0.28	0.313333	0.65	1.174603	1.35

图 12-25 包头一区土壤单一重金属评价分析结果导出

选择"区域评价模型"子菜单,执行土壤重金属神经网络综合评价模型评价分析,分析结果如图 12-26 所示。

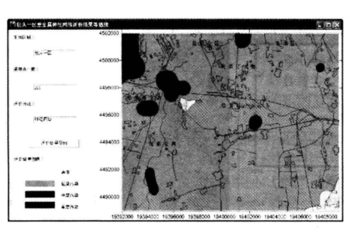


图 12-26　包头一区土壤重金属神经网络综合评价分析结果

选择“国家标准综合评价模型”子菜单，执行土壤重金属国家标准综合评价模型评价分析，分析结果如图 12-27 所示。

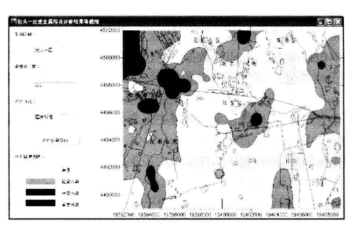


图 12-27　包头一区土壤重金属国家标准综合评价分析结果

12.4.3　预测分析模型

选择“预测分析模型”子菜单，执行采样区域 5 年、10 年、15 年、20 年后土壤重金属污染情况的分析，分析结果如图 12-28 所示。按照国家标准，将污染情况划分为四个等级，并绘制成等值线。

如果在预测分析模型参数设置区域修改了模型的某个参数，单击“修改参数并重新预测”按钮，然后再次选择“预测分析模型”子菜单，执行新的预测分析。

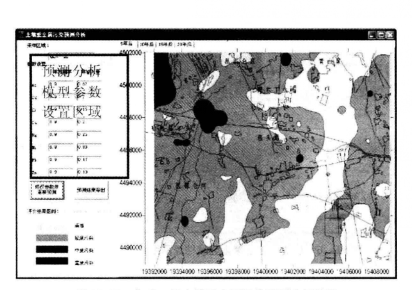


图 12-28　包头一区土壤重金属污染预测分析结果

12.5　数据管理

数据管理菜单包括植物数据、土样数据、大气数据和水数据，实现数据库中数据的更新。“植物数据”子菜单包括：植物采样点数据、重金属数据、光谱数据和数据库中数据的导入导出。

打开某一工作空间后，可以把该采样区域中采集的所有植物样本实验室测得的重金属含量数据从数据库中一次导出到 Excel。

打开某一工作空间后，可以把该采样区域中采集的所有植物样本实验室测得的光谱数据从数据库中一次导出到 Excel。

“土样数据”子菜单包括：土壤采样点数据、土壤形态数据及土壤重金属数据和数据库中数据的导入导出。

打开某一工作空间后，可以把该采样区域中采集的所有土壤样本采样点详细信息数据从数据库中一次导出到 Excel。

打开某一工作空间后，可以把该采样区域中采集的所有土壤样本实验室测得的形态数据从数据库中一次导出到 Excel。

打开某一工作空间后，可以把该采样区域中采集的所有土壤样本实验室测得的重金属含量数据从数据库中一次导出到 Excel。

“大气数据”子菜单实现典型区域采集的大气数据和数据库中数据的导入导出，操作方法与植物数据的导入导出类似。

“水数据”子菜单实现典型区域采集的水样数据和数据库中数据的导入导出，操作方法与植物数据的导入导出类似。

系统运行时会生成一些临时文件，保存在……\project\temp 文件夹中，主要是对数据分析形成的饼图、柱状图、等值线图等临时文件的存储。可以把该文件夹中的临时文件删除。

参考文献

[1]Alloway B. J. Heavy Metals in Soils (2nd Edition) [M]. Publisher of Springer, 31 Dec 1994: 230-238.

[2]Beni C, Liu X. M. A Least Biased Fuzzy Clustering Method [C]. IEEE Trans of PAMI, 1992, 16(9):954-960.

[3]Bezdek J C. Clustering Validity with Fuzzy Sets [J]. Mathematical Biology, 1974, 1:57-71.

[4]Bezdek J C. Pattern Recognition with Fuzzy Objective Function Algorithms [M]. Plenum Press, New York, 1981:78-81.

[5]Swaine D J. The Trace Element Content of Soils [J]. Commonwealth Bur. Soil Science. Harpenden of Tech. Commune. 1955(48): 157-170.

[6]Ester M, Kriegel H. P, Sander J, et al. A Density BP Algorithm for Discovering Clusters in Large Spatial Data with Noise [C]. The 2nd the International Conference on Knowledge Discovery and Data Mining, Portland, 1996:348-356.

[7] Florian Statescu, Dorin Cotiusca Zauca. Heavy Metal Soil Contamination [J]. Environmental Engineering and Management Journal, 2006, 5(5):1205-1213.

[8]Guha S, Ratogi R, Shim K. CURE: An Efficient Clustering Algorithm for Large Data [C]. 1998 ACM-SIGMOD International Conference on Management of Data (SIGMOD 98), Seattle,Washington, 1998:2567-2578.

[9]Hagen L, Kahng A B. New Spectral Methods for Ratio Cut Partitioning and Clustering [J]. IEEE Trans Computer-Aided Design, 1992, 11(9): 1074-1085.

[10] Hamad D, Biela P. Introduction to Spectral Clustering [C]. Proceedings of 3rd International Conference on Information and Communication Technologies: From Theory to Applications. New York, 2008: 1-6.

[11]Han J. W., Kamberm. 范明,孟小峰,译. Data Mining Concept and Techniques [M]. 北京:机械工业出版社, 2001.

[12]Hinneburg A, Keim D A. An Efficient Approach to Clustering in Large Multimedia Data with Noise [C]. The 1998 International Conference on Knowledge Discovery and Data Mining (KDD 98), New York, 1998.

[13]Jain A,Murty M, Flynn P. Data Clustering: A Review [J] . ACM Computing Surveys, 1999, 31(3): 2642-2323.

[14]Kang-tsung Chang 著,陈健飞译.地理信息系统导论[M].北京:科学出版社,2004.

[15]Kaufman L, Rousseeuw P J. Finding Groups in Data: An Introduction to Cluster Analysis

[M]. New York: John Wiley & Sons, 1990.
[16] Macqueen J. Some Methods for Classification and Analysis of Multivariate Observations [C]. The 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 1967.
[17] Shi J, Malik J. Normalized Cuts and Image Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000, 22(8): 888-905.
[18] Xiang Li, Lingling Li. The Smart System of Soil Heavy Metal Pollution Evaluation and Safety Warning Using GIS and MAS [J]. Journal of Environmental Research and Development, 2011, Volumes 6(1): 90-98.
[19] Xiang Li, Xiaoyu Zhang. Evaluating Multiple Heavy Metal Pollutants in Soil by Artificial Neural Network: A Case Study in Baotou China [C]. Energy Procedia, 2011 Volume 11: 4627-4631.
[20] Xiang Li. Improved Soil Heavy Metal Pollution Evaluating System Using MAS and GIS [C]. Energy Procedia, 2011 Volume 11: 4242-4246.
[21] Zahn C T. Graph the Oriented Methods for Detecting and Describing Gestalt Clusters [J]. IEEE Transaction on Computers, 1971, 20: 68 -86.
[22] Zhang T, Ramakrishnan R, Livny M. Birch. An Efficient Data Clustering Method for Very Large Data [C]. The International Conference Management of Data, Montreal, Canada, 1996.
[23] 卜德云,张道强. 自适应谱聚类算法研究[J]. 山东大学学报(工学版),2009,39(5):22-26.
[24] 卜德云. 自适应谱聚类算法的研究与应用[D]. 南京航空航天大学硕士学位论文,2010.
[25] 蔡晓妍,戴冠中,杨黎斌. 谱聚类算法综述[J]. 计算机科学,2008,35(7):14-18.
[26] 戴月明,高倩. 自适应半监督模糊谱聚类算法[J]. 计算机工程与应用,2010,46(33):212-214.
[27] 冯春雨,白红娟,肖根林,等. 重金属污染土壤的生物修复研究现状[J]. 工业安全与环保,2010,36(4):27-28.
[28] 高新波. 模糊聚类分析及其应用[M]. 西安电子科技大学出版社,2004:49-61.
[29] 葛哲学,孙志强. 神经网络理论与 MATLABR2007 实现[M]. 北京:电子工业出版社,2007.
[30] 龚健雅. 当代 GIS 若干理论和技术[M]. 武汉:武汉测绘科技大学出版社,1999.
[31] 胡大伟,卞新民,李思米,等. 基于神经网络的农田土壤重金属空间分布分析[J]. 农业环境科学学报,2007(1):238-243.
[32] 胡守仁. 神经网络应用技术[M]. 北京:国防科技大学出版社,1993.
[33] 贾建华. 谱聚类集成算法研究[M]. 天津:天津大学出版社,2011.
[34] 黎夏,叶嘉安,刘小平,等. 地理模拟系统——元胞自动机与多智能体[M]. 北京:科学出版社,2007.

[35] 李德仁,王树良,李德毅,等. 论空间数据挖掘和知识发现的理论和方法[J]. 武汉大学学报(信息科学版),2002,27(3):221-233.

[36] 李斐,王殊伟,柯宝贵. 应用聚类分析方法进行实测重力数据的选点优化[J]. 武汉大学学报(信息科学版),2009,24(3):257-261.

[37] 李军,孙春宝,李云,等. 我国大气铅浓度水平与污染源排放特征[J]. 化工环保,2009,29(4):376-380.

[38] 李向,管涛,徐清. 基于 BP 神经网络的土壤重金属污染评价方法——以包头土壤环境质量评价为例[J]. 中国农学通报,2012,28(2):250-256.

[39] 李向,宋涛. MAS 技术在土壤重金属污染评价及预警中的应用[J]. 计算机技术与发展,2010,20(1):217-220.

[40] 李向,徐清. 基于灰色关联分析理论的典型区域土壤重金属污染评价研究[J]. 安全与环境学报,2012,(12)1: 150-154.

[41] 李泽琴,侯佳渝,王奖臻. 矿山环境土壤重金属污染潜在生态风险评价模型探讨[J]. 地球科学进展,2008,23(5):509-516.

[42] 李战,李坤. 重金属污染的危害与修复[J]. 现代农业科技. 2010,16:268-270.

[43] 李志霞. 环境监测-理论篇[M]. 2 版. 大连:大连理工大学出版社,2010 年 4 月.

[44] 凌辉,武伟,王润,等. 小尺度下土壤重金属铬含量的空间插值方法比较研究[J]. 西南大学学报(自然科学版), 2007(11):379-342.

[45] 刘春阳,张宇峰,滕洁. 土壤中重金属污染的研究进展[J]. 污染防治技术,2006,19(4):42-45.

[46] 刘梅,李发鹏,卢善龙. 流域系统重金属污染研究进展[J]. 安徽农业科学,2011,39(25):15622-15626.

[47] 马溪平,李法云,肖鹏飞,等. 典型工业区周围土壤重金属污染评价及空间分布[J]. 哈尔滨工业大学学报,2007,39(2):326-329.

[48] 孟宪丽. GIS 支持下的吉林省黑土区土壤重金属污染评价研究[D]. 东北师范大学,2008.

[49] 潘锋,王建东,牛奔. 基于谱分析的无监督特征选择算法[J]. 计算机应用,2011,31(8):2108-2110.

[50] 施培蓓,郭玉堂,胡玉娟,等. 多尺度的谱聚类算法[J]. 计算机工程与应用,2011,47(8):128-130.

[51] 檀满枝,陈杰,徐方明,等. 基于模糊集理论的土壤重金属污染空间预测[J]. 土壤学报,2006,43(3):389-396.

[52] 檀满枝,陈杰,郑海龙,等. 模糊 c-均值聚类法在土壤重金属污染空间预测中的应用[J]. 环境科学学报,2006,26(12):2086-2092.

[53] 中华人民共和国环境部. 土壤环境监测技术规范[M]. 中华人民共和国环境保护行业标准 HT/T166-2004. 2004,12.

[54] 王国庆,骆永明,宋静,等. 土壤环境质量指导值与标准研究——国际动态及中国的修订考虑[J]. 土壤学报,2005,4:23-36.

[55]王海军,张德礼. 基于空间聚类的城镇土地定级方法研究[J]. 武汉大学学报(信息科学版),2006,3(7):628-631.
[56]王玲,薄列峰,焦李成. 密度敏感的半监督谱聚类[J]. 软件学报. 2007,18(10):2412-2422.
[57]吴健生,宋静,郑茂坤,等. 土壤重金属全量监测方法研究进展[J]. 东北农业大学学报,2011,42(5):133-139.
[58]谢正苗,李静,王碧玲,等. 基于地统计学和 GIS 的土壤和蔬菜重金属的环境质量评价[J]. 环境科学,2006(27)10:246-251.
[59]徐礼生,吴龙华,高贵珍,等. 重金属污染土壤的植物修复及其机理研究进展[J]. 2010(38)3:372-377.
[60]许嘉林,杨居荣. 陆地生态系统中的重金属[M]. 北京:中国环境科学出版社,1995.
[61]姚申君,吴健平,易敏,等. GIS 在环境影响评价中的应用[J]. 环境科学导刊,2007,26(6):77-80.
[62]赵其国,郑茂坤. 未来土壤学一书全文翻译[J]. 土壤,2007,39(1):94-152.
[63]郑海龙,陈杰,邓文靖,等. 城市边缘带土壤重金属空间变异及其污染评价[J]. 土壤学报,2006,43(1):39-45.
[64]郑向群,赵政,刘东生. 基于数据仓库的土壤环境监测综合挖掘模型构架[J]. 农业工程学报,2008,24(8):162-168.
[65]周脚跟,赵春江. 基于局部离群指数的土壤重金属污染评价方法[J]. 农业工程学报,2010,26(1):279-283.
[66]周生路. 土地评价学[M]. 南京:东南大学出版社,2006.